ANALYSIS OF THE SEQUENCE ORGANISATION OF PART OF THE RIGHT ARM OF THE THIRD CHROMOSOME OF *DROSOPHILA MELANOGASTER*

John S. de Banzie

Doctor of Philosophy

University of Edinburgh

1981

ABSTRACT.

The sequence organisation of a long continuous segment of a particular region of the genome can be studied by the isolation of recombinants carrying overlapping fragments of DNA from that region. This technique is known as "walking along the chromosome". The advantages and disadvantages of this approach are discussed. In this investigation the technique was used to isolate DNA sequences from part of the right arm of the third chromosome of Drosophila melanogaster strains Canton S and Oregon R (region 84 of the salivary gland chromosome map (Bridges, 1935)), using libraries of cloned embryo DNA.

In three screenings of the libraries, twenty seven different recombinants were identified, of which thirteen were chosen for further study. Of these, nine carry overlapping fragments from two separate 35kb segments of the 84 region, and four come from other parts of the genome, but have homology with short repeated sequences present on some of the other recombinants. Evidence is presented that in three of these cases the short repeats are tRNA genes. Recombinants were characterised by restriction mapping and localisation of repetitive sequences. A repetitive sequence present at a particular site in the Canton S genome was found to be missing from the corresponding position in the Oregon R genome. The significance of this finding is discussed.

ACKNOWLEDGEMENTS.

I would like to thank:

for specific things	-	J. Dahlberg and E. Lund for labelled 8 to 4S RNA W. Gehring, S. Artavanis-Tsakonas and P. Schedl for 14C4 D. Hogness for the OR library (via M. Wolfner and D. Glover) T. Maniatis for the CS library (via J. Lauer) M. Mitchell for typing (and coffee) D. Sherrat for his restriction map of RSF2124
for lots of things	-	Alan, Anne, Binie, Douglas, Graham, Joan, Joey, Karen, Kathy, Les, Marion, Mike, Mitch, Noreen, Pam, Pete, Peter, Ruby and Sandra
especially	-	Barbara Will and Sue Dunbar
and above all	-	David Finnegan

CONTENTS.

ABBREVIATIONS.

DNase = deoxyribonuclease
kb = kilobases
RNase = ribonuclease
ul = microlitre

All other abbreviations in the text are according to British Standard BS 3763:1976 "The International System of units (SI)".

Journal titles are abbreviated following the standard adopted by the American Chemical Society.

Section 1.

INTRODUCTION.

The central problem of molecular biology at this time is the means by which growth and differentiation are achieved in higher organisms. Fundamentally this is a question of the regulation of gene expression. It seems likely that the means by which regulation is achieved will impose certain constraints on the structural and sequence organisation of the genome.

Structural organisation of the genome.

The smallest unit of packaging of DNA in chromatin in the nucleosome; this consists of about 200 base pairs of DNA wound around a protein core consisting of two molecules each of histones H2a, H2b, H3 and H4 (see, for example, Felsenfeld, 1978; Klug et al, 1980). The next level of structure, known as a solenoid, is formed by coiling of a filament of nucleosomes into a helix of approximate diameter 30nm. Histone H1 appears to be involved in stabilisation of this structure (Finch and Klug, 1976). A detailed model of chromatin structure incorporating these features is described by Worcel and Benyajati (1977). Finally, in interphase chromosomes, strands of chromatin appear to be cross-linked, giving rise to large, independently supercoiled loops (Benyajati and Worcel, 1976; Igo-Kemenes and Zachau, 1977) known as domains.

The structural organisation of specific sequences in chromatin can be examined using nuclease digestion. It has been shown that the globin (Weintraub and Groudine, 1976) and ovalbumin genes (Palmiter et al, 1977) are more sensitive to digestion with DNase 1 in tissues in which they are expressed than in other tissues. Similarly, the heat shock genes of Drosophila melanogaster show increased sensitivity to DNase 1 on activation (Wu et al, 1979b; Wu, 1980). Thus gene activation appears to involve a degree of "unpackaging" of the transcribed sequences, rendering them more accessible to nuclease attack.

Sites of exceptional sensitivity to DNase 1 digestion have been found at the 5'ends of non-induced heat shock genes in all cases so far examined (Wu et al, 1979a; Wu, 1980). The site closest to the 5'end of the transcribed region lies in, or close to, a sequence common to several heat shock genes (Moran et al, 1979; Ashburner and Bonner, 1979). Therefore it has been suggested that this site represents a regulatory sequence, which is relatively exposed in order to allow interaction with an inducer of the heat shock genes (Wu, 1980). On binding of the inducer, the heat shock genes are unpackaged and thus become available for transcription.

Sequence organisation of the genome.

Many models for the regulation of gene expression have been proposed which involve repetitive sequences (Britten and Davidson, 1969; Georgiev, 1969; Paul, 1972; Vaughan, 1977; Davidson and Britten, 1979). These models all predict that repetitive sequences will be found in a particular relationship to transcribed sequences.

In the majority of higher organisms (see Lee et al, 1977 for list and references) between 35 and 70% of the genome consists of repetitive sequences of length 0.2 - 0.5kb alternating with unique sequences of length 1 - 3kb. This is known as the short period interspersion pattern (Lee et al, 1977). In Drosophila melanogaster (Manning et al, 1975; Crain et al, 1976), Chironomus tentans (Wells et al, 1976) and Caenrhabditis elegans (Schachat et al, 1978), DNA of this pattern is not detectable, the interspersed repetitive and unique sequences being much longer.

A few repetitive sequences have been studied in detail: satellite DNAs (see Lewin, 1974), rRNA genes (see, for example, Glover and Hogness, 1977; Wellauer and Reeder, 1975; Cory and Adams, 1977), histone genes (Kedes, 1979), and 5S genes (Brown et al, 1971; Artivanis-Tsakonas et al, 1977; Hershey et al, 1977) which are tandemly repeated; and genes of the 412 type (Finnegan et al, 1977; Strobel et al, 1979) which are dispersed throughout the genome.

The majority of moderately repetitive sequences in Drosophila melanogaster can be assigned to discrete, conserved families of dispersed repeats, (Wensink, 1977; Young, 1979; Wensink et al, 1979), the locations of which vary from strain to strain (Strobel et al, 1979; Young, 1979; Yen and Davidson, 1980; W. Bender and P. Spierer, personal communication; B. Will and D. Finnegan, personal communication; see also section 4). These repeats are most often found in long blocks (Manning et al, 1976). Two forms of organisation have been demonstrated for these blocks of sequence. In the first, typified by the 412 gene family (Finnegan et al, 1977), the entire sequence is repeated intact at all sites at which it occurs, with little variation. In the second, typified by the sequences described by Wensink et al (1979), the repetitive block is composed of several much shorter repeated elements. Different repetitive blocks of this type may share several short repeated elements, but in a scrambled arrangement.

I propose that all repetitive elements are originally of the 412 type, and that over a long period, through transposition within the genome (as evidenced by the different locations of middle repetitive sequence in different strains (see above)), combined with occasional defective excision, these elements become scrambled with other sequences, until their original form is no longer recognisable. Thus the 412 family would be an example of a sequence which arose relatively recently, and has not yet been scrambled, and the scrambled clusters of Wensink et al(1979) the remnants of much older sequences. Alternatively, there could be some selective pressure on the 412 gene family which maintains it in its current form. Such a model for the evolution of repetitive sequences in Drosophila would predict a continuum of such sequences having degrees of scrambling between that of 412 and that of the Wensink clusters. There is some evidence for differing degrees of heterogeneity within different

families of repetitive sequence (D. Finnegan, personal communication; G. Rubin, personal communication; V. Pirrotta, personal communication). Note that the amount of variation required to render members of a repeated family unrecognisable as such will be fairly limited.

The best study of the relationship between a specific structural gene and adjacent repetitive sequences is that of Shen and Maniatis (1980), involving a cluster of four β-like globin genes in rabbit. The results were complex, but each gene was flanked by at least one set of short (0.14 to 0.4kb) inverted repeats, and the entire cluster by a 1.4kb inverted repeat. However, given that this gene cluster may well have evolved by gene duplication, and in the absence of comparable examples, the significance of these findings is unclear.

The majority of the information set out above has been obtained by the application of recombinant DNA technology. This enables specific fragments of the genome to be prepared in quantities sufficient for analysis. One way in which this technology may be exploited is known as "walking along the chromosome". Using this technique, a series of recombinants are isolated which carry overlapping fragments of DNA representing a long continuous stretch of the genome. These recombinants can then be used to determine the sequence organisation of this particular segment of the genome.

"Walking along the chromosome"

In order to "walk along the chromosome" a collection of cloned fragments of genomic DNA representing the entire genome of the organism is needed. Such a collection is known as a library. The library is screened with a probe sequence which has homology with the region in which the "walk"is to take place. This sequence must be unique or recombinants carrying fragments from other regions will be picked up. The recombinants isolated in the screening are characterised with respect to the extent and position of overlaps with the probe and any repetitive sequences located. A restriction fragment carrying only unique sequence and lying as far from the original probe sequence as possible is then purified and used as probe in a second screening of the library. This cycle of probe-characterise-purify-probe (illustrated in figure 1.1) is repeated until the objective of the "walk" is attained. Each feature of this process will be dealt with in detail below.

Libraries

There are three classes of vector which may be used to construct a library. These are cosmids (Collins and Hohn, 1978), phage, and plasmids. Phage are by far the best vector systems for "walking", combining moderate length inserts with simple and rapid screening by plaque hybridisation (Benton and Davis, 1977). Cosmids and plasmids must be screened by colony hybridisation (Grunstein and Hogness, 1975) which is less convenient. The large inserts of cosmids will result in more rapid progress per screening, but this will tend to be offset by the difficulty in characterising such a long segment of DNA. However cosmids will be particularly useful where a long stretch of repetitive sequence has to be bridged (see below).

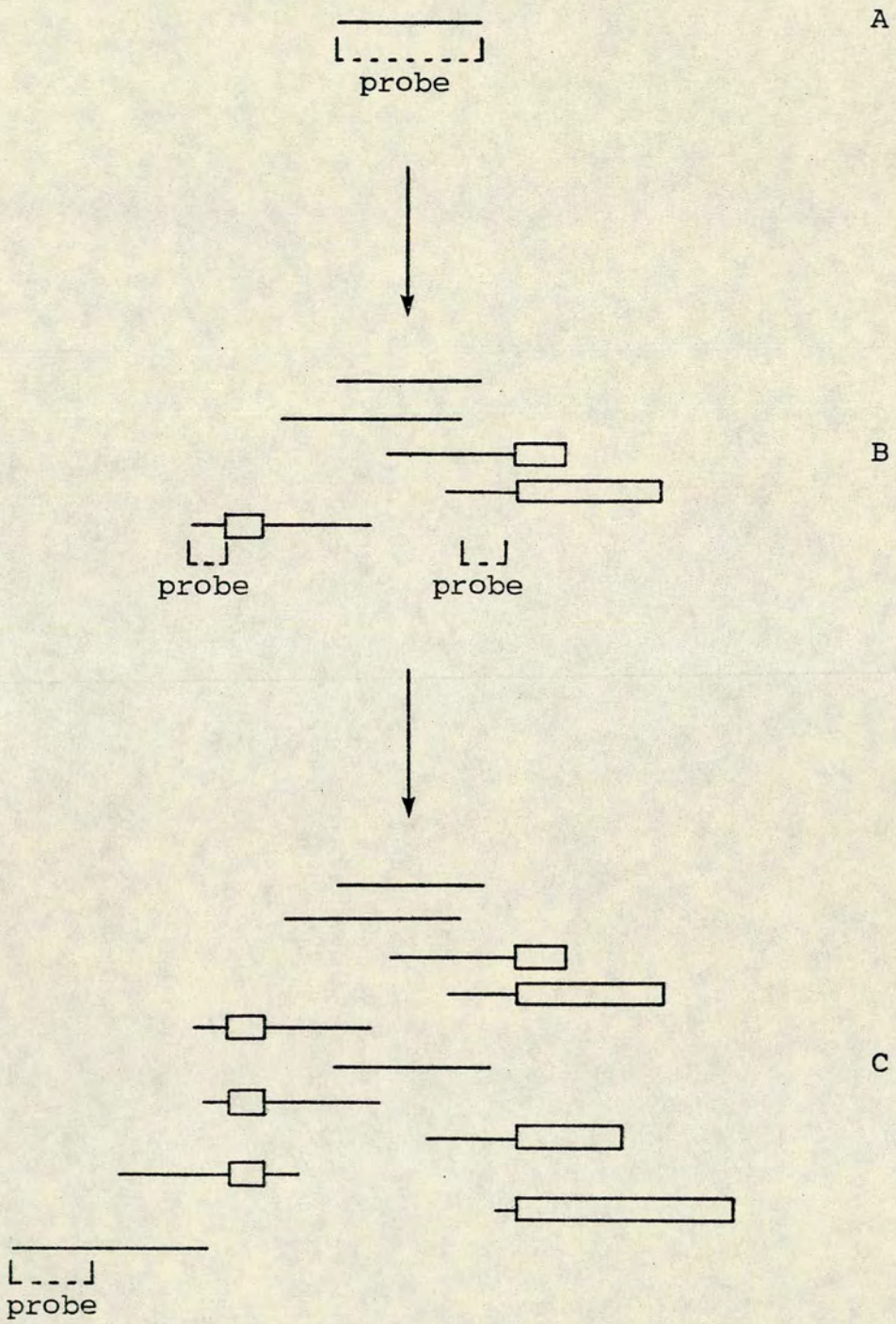
It is important that the inserts are generated with random ends. The best way to do this is to shear the DNA mechanically. If the ends are not random, as would be the

Figure 1.1
Diagram illustrating the process of "walking along the chromosome".

The walk is started using a probe (A) derived from the region of interest, and shown to be composed entirely of non-repetitive sequences. (Repetitive sequences are boxed). This probe is used to screen a library, and recombinants carrying overlapping fragments isolated and characterised (B). Fragments of these recombinants, carrying only unique sequences, and lying as far from the original probe sequence (A) as possible, are purified and used in a second screening of the library. A second generation of overlapping recombinants is isolated, characterised, and fresh probe fragments prepared for a third screening (C).

Note that at the right end of the "walk" illustrated here, a long region of repetitive sequence has been encountered. This repetitive region is too long to be spanned by the inserts of the library being used. Thus the "walk" is blocked in this direction as no unique sequence beyond the repetitive sequence can be isolated. It may be possible to resume progress in this direction by (1) switching to a library carrying longer inserts (for example, a cosmid library) or (2) switching to a library prepared from DNA of a different strain (see text).

Figure 1.1



case if they were produced by partial digestion with a restriction enzyme, it is likely that some sequences will not be represented in the library. For example, long stretches of DNA having no suitable restriction sites could not be cloned, and regions where a large number of restriction sites lie close together would at best be under-represented. Such situations can occur in satellite DNA and other tandemly repeated sequences.

Inserts generated by shearing can be joined to vector DNA in three ways: blunt end ligation (Sgaramella and Khorana, 1972), oligo dC:dG (or dA:dT) tailing (Wensink et al, 1974), or use of linkers (Maniatis et al, 1978). Assuming that a phage vector is to be used, the best method is to use linkers. This enables the insert to be cleaved from the vector DNA in its entirety which is a great advantage when the recombinants are being characterised.

The method chosen to link inserts to vector DNA will determine the frequency of double inserts; that is, an insert generated by ligation of two fragments of DNA from separate regions of the genome. Such inserts could result in the "walk" being transposed to another region of the genome. Direct ligation will produce double inserts at a frequency dependent on the ratio of vector to insert DNA molecules, and, where the insert size is limited, as in the case of cosmid and phage vectors, on the distribution of insert sizes. Hence with cosmid and phage vectors double inserts can largely be avoided by isolating the appropriate size class of insert molecules prior to ligation. The same arguments apply when ligation is with linkers, but as there are two ligation steps the frequency of double inserts will be higher, although this can be countered by having a high ratio of linker to insert molecules in the first ligation step.

Double inserts can be avoided entirely by using tailing to link vector and insert. However this prevents the insert from being cleaved from the vector, with concomitant difficulties in characterisation of the

recombinants (an exception to this is where inserts are dC:dG tailed into a Pst 1 site, but this is not yet possible with phage vectors). Thus, provided the inserts are sized prior to ligation, the use of linkers represents the best method of constructing a library for "walking".

The source of DNA used in constructing the library must also be considered. It must always be borne in mind that a cloned sequence is derived from one chromosome of one cell of one individual, and therefore may not be representative of the "normal" sequence. In addition it is possible that rearrangements of sequence may take place during development and differentiation, as for example with the immunoglobulin genes and the genes determining mating type in yeast (Hozumi and Tonegawa, 1976; Hicks et al, 1979). Both these possibilities can be tested by probing gel transfer filters of restricted bulk DNA prepared from different tissues or developmental stages with the recombinant in question, although the interpretation of the results of such experiments is difficult (see below and section 4.4). Thus, unless one is particularly interested in the sequence organisation of a particular tissue or developmental stage, it is best to prepare DNA from material having relatively few differentiated cell types. In Drosophila, embryos represent the most useful source of DNA.

Having obtained a library it is likely to be necessary to amplify it. This will however reduce the efficiency of the screening experiments. Consider the proportion of the genome represented in a screening experiment.

Let a be the average length of insert in the recombinant library in kb.
G be the haploid genome size in kb.
N be the number of independent recombinant phage recovered when the library was first manufactured.
n be the number of phage screened.
p be the proportion of the library amplified.

The proportion of the genome represented by n plaques of the unamplified library is given by

$$1 - (1 - a/G)^n \qquad \text{(Clarke and Carbon, 1976)} \qquad (1.1)$$

After the first round of amplification of the entire library (p = 1), assuming all recombinants to be amplified equally (see below), the proportion of the library represented by n plaques is given by

$$1 - (1 - 1/N)^n \qquad (1.2)$$

And hence the number of different recombinants present in n plaques of the amplified library is

$$(1 - (1 - 1/N)^n)N \qquad (1.3)$$

Thus the proportion of the genome represented by n plaques of the amplified library is

$$1 - (1 - a/G)^{((1 - (1 - 1/N)^n)N)} \qquad (1.4)$$

If only a fraction of the library is amplified this expression becomes

$$1 - (1 - a/G)^{((1 - (1 - 1/Np)^n)Np)} \qquad (1.5)$$

Thus the proportion of the genome represented in a given number of plaques is reduced on amplification of the library. Subsequent rounds of amplification will compound this effect. However, note that when Np is very much larger than n the effect will be minimal.

So far it has been assumed that on amplification all recombinants are amplified equally. However, not all phage will be amplified to the same extent, even if they are all of equal viability (see below). Rather, the numbers of the different recombinants after amplification will follow a normal distribution about some mean value. Subsequent rounds of amplification will tend to favour those recombinants present in greater numbers and these will form an increasingly large proportion of the library. Similarly, recombinants which were under-amplified in the first round will tend to form a still smaller proportion of the library, and may eventually be lost altogether.

The rate with which this effect will make itself felt is not known, but it seems likely that several rounds of amplification would be required before it became significant. However, this is dependent on the fraction of the library amplified; the smaller the fraction, the sooner the effect becomes significant.

If, due to the different inserts carried, recombinants are not of equal viability, the effects are obvious. Phage which grow poorly will be under-amplified, and phage which grow well, over-amplified. The former might eventually be lost from the library. The viability of a recombinant might be affected by the length of the insert which it carries. Phage particles which carry short DNA molecules tend to be unstable, and will therefore be under-represented in the amplified library.

Viability may also be affected by some aspect of the sequence of the insert. For example, if part of the insert were to be transcribed (and possibly translated) during phage growth, the sequestration of resources which might otherwise be used for phage production might reduce the viability of that recombinant. The viability of a recombinant might also be increased by its insert. Phage which are *red*$^-$ *gam*$^-$ must carry short sequences known as *chi* elements if they are to package replicated DNA into mature phage particles (Stahl, 1979). If the insert of a recombinant contained a *chi* element, this might increase the number of mature phage produced. Note that where the viability is affected by the sequence of the insert all recombinants carrying that sequence could be affected. This might result in some regions of the genome being over- or under-represented in the library.

In summary, amplification will, for various reasons, result in a reduction in the number of different recombinants present in a given number of plaques. Hence the efficiency of screening will be reduced. This may be minimised by the following

(1) Never amplify a library unless absolutely necessary.

(2) When amplifying, amplify as large a fraction of the library as possible.

(3) When amplifying, grow the phage at a lower density than usual, say 10 000 plaques per plate. This will minimise the effects of differential viability by reducing competition between the different phage.

Probes

The only requirements of the probes used in "walking" are that they consist entirely of unique sequence, do not cross-hybridise with either vector or E. coli DNA, and, of course, are derived wholly from the region of the genome in which the "walk" is taking place.

If a probe containing repetitive sequences were used recombinants carrying inserts from many regions of the genome would be isolated (unless the sequence were tandemly repeated at only one site). At best this would delay the "walk", but under certain circumstances, where a low frequency repeat lay at the end of a probe, it might result in the "walk" being transposed to another region of the genome entirely. The identification of repetitive sequences is therefore one of the most important aspects of carrying out a successful "walk". There are two methods by which repetitive sequences can be usefully identified.

In the first, and as far as "walking" is concerned, more useful method, the DNA to be tested for repeated sequences is digested with a restriction enzyme (or enzymes) and the fragments separated by electrophoresis through an agarose gel. The fragments are then transferred to a nitrocellulose filter and probed with total genomic DNA (preferably from the same source as that used in constructing the library) which has been nick translated to a high specific activity. If a fragment of the DNA being tested carries only single copy sequence, it will hybridise with this sequence in the probe, and, on autoradiography, give a band the intensity of which will be proportional to its length. If a fragment carries repetitive sequence it will also hybridise the repeats present in the probe DNA, and, on autoradiography, give a correspondingly more intense band. The increased intensity will be related to both the copy number and length of the repeated sequence, provided

that the hybridisation conditions are such that saturation of the filter bound DNA does not occur. In order for a fragment to be reliably detected as carrying repetitive sequence, the product of the copy number and the length of the repeated sequence will need to be about twice the length of the fragment, thus giving about three times the intensity of hybridisation expected for a band of that size. Hence it is best to choose restriction enzymes which cut the DNA fairly frequently. This method allows one to locate a repetitive sequence with respect to the restriction map of the DNA involved. A large number of sequences can be screened rapidly and economically (only one nick translation is required) but little information is gained about the repeated sequence itself.

The second method gives much more information, but is much less convenient. Total genomic DNA, ideally from the same tissue and developmental stage as that used in constructing the library, is digested with one or more restriction enzymes. The fragments are electrophoresed through an agarose gel and transferred to nitrocellulose. The nitrocellulose filter is then probed with the nick translated DNA which is to be tested for repetition. The prediction is that if the sequence is unique, the number of fragments hybridising in the total genomic DNA will be the same as the number of segments of genomic sequence produced by digestion of the probe. If the probe contains repetitive sequence a greater number of fragments will hybridise in the genomic DNA (with one exception which will be discussed below). The sensitivity of this method is limited by the amount of total genomic DNA which can be loaded onto a gel, and the specific activity of the probe. Under the conditions used in this investigation repeats of length greater than 0.2kb would be detected.

The problem with this method is that extra bands may appear for reasons other than repetition. If the total genomic DNA is not digested to completion, or if the enzyme used has a contaminating activity (for example, Hind 111 may be contaminated with Hind 11), extra bands will be

present on the autoradiograph. It is possible to control against this particular difficulty by running two tracks of total genomic DNA, one containing DNA digested with an amount of enzyme expected to give total digestion, the other DNA digested with twice that amount of enzyme. If the two tracks are not identical on the autoradiograph, then under digestion or a contaminating activity is a problem and the results must be interpreted with great care.

Other sources of extra bands are heterogeneity of restriction sites within the population, probing with a recombinant containing a double insert, probing with a recombinant which has undergone a rearrangement during the cloning procedure (which is really a specialised case of a double insert), or probing with cDNA or mRNA derived from a gene with introns. Thus, given that there are several explanations other than repetition for a small number of extra bands, and that the frequency of double inserts will be very low in a properly constructed library, it is suggested that a fragment which gives rise to one or two extra bands may still be used as a probe in screening the library. Provided that the analysis of the recombinants isolated is thorough, transposition of the "walk" due to low frequency repeats (or double inserts) will be detected at that stage.

As described so far, this method would not detect as repetitive a probe the entire sequence of which was repeated as an intact unit. Such a probe would give the same number of hybridising fragments as a unique sequence, but the level of hybridisation would be increased in proportion to the frequency of repetition. Therefore a track should be included on the gel to show the intensity of hybridisation expected for single copy sequence. The simplest and most useful way to do this is to include a track containing an amount of unlabelled digested probe DNA equivalent to single copy in the total genomic DNA tracks. However it is unlikely that repeats occurring in less than 3 to 5 copies would be detected.

Now consider what other information may be derived from these results. If the probe is unique, and has ends which were not generated by the same restriction enzyme as used in the experiment, the distance to the nearest restriction sites beyond the sequence used as probe can be determined. This information can be helpful in restriction mapping and determining overlaps of isolated recombinants. If the probe is repetitive, then the number of extra bands observed indicates the minimum number of copies of the repeat (or repeats). The intensity of the bands gives an indication of the maximum length of the repeats. If a particular fragment corresponding in size with a fragment of the probe is labelled more heavily than would be expected from its length, then it is likely that the sequence present on that entire fragment of the probe is repeated as an intact unit. Some possible results are illustrated in figure 1.2.

In conclusion, the first method is better for the identification of a repeated sequence, and the second for characterisation of that sequence. Note that neither method will reliably identify low frequency repeats, those most likely to transpose the "walk" to another region of the genome. This will be discussed further when the means of determining regions of overlap between recombinants is considered.

Suppose it is found that there is a long repetitive sequence in the path of the "walk" (as at the right hand end in figure 1.1) such that on probing with the most extreme unique sequence available, none of the recombinants isolated span the repeated region. Unless a unique sequence from the far side of the repeat can be obtained the way forward is blocked. There are two ways round this problem. One is to switch from using a phage library to a cosmid library and hope that the large inserts carried by cosmids will be long enough to span the repetitive region. Alternatively, at least in Drosophila melanogaster, one can switch to a library constructed with DNA from a different strain. In no case so far examined at this level

Figure 1.2
Diagram illustrating some possible results on probing restriction enzyme digested total genomic DNA with a cloned fragment.

The upper line represents the cloned fragment used as probe. The next line or group of lines represent the regions of the genome having homology with the cloned fragment, including the sequence from which the cloned fragment is derived. The illustrations at the bottom represent the autoradiographs obtained on digesting probe (left) and genomic (right) DNA with the restriction enzyme, separating the fragments by agarose gel electrophoresis, transferring to nitrocellulose, and hybridising with the probe. The probe (left) track indicates the intensity expected for single copy sequences. Boxed segments represent repetitive sequences, arrowheads sites for the restriction enzyme.

Situation A - The cloned fragment used as probe carries only unique sequence. It labels three fragments: one internal fragment and two flanking fragments.

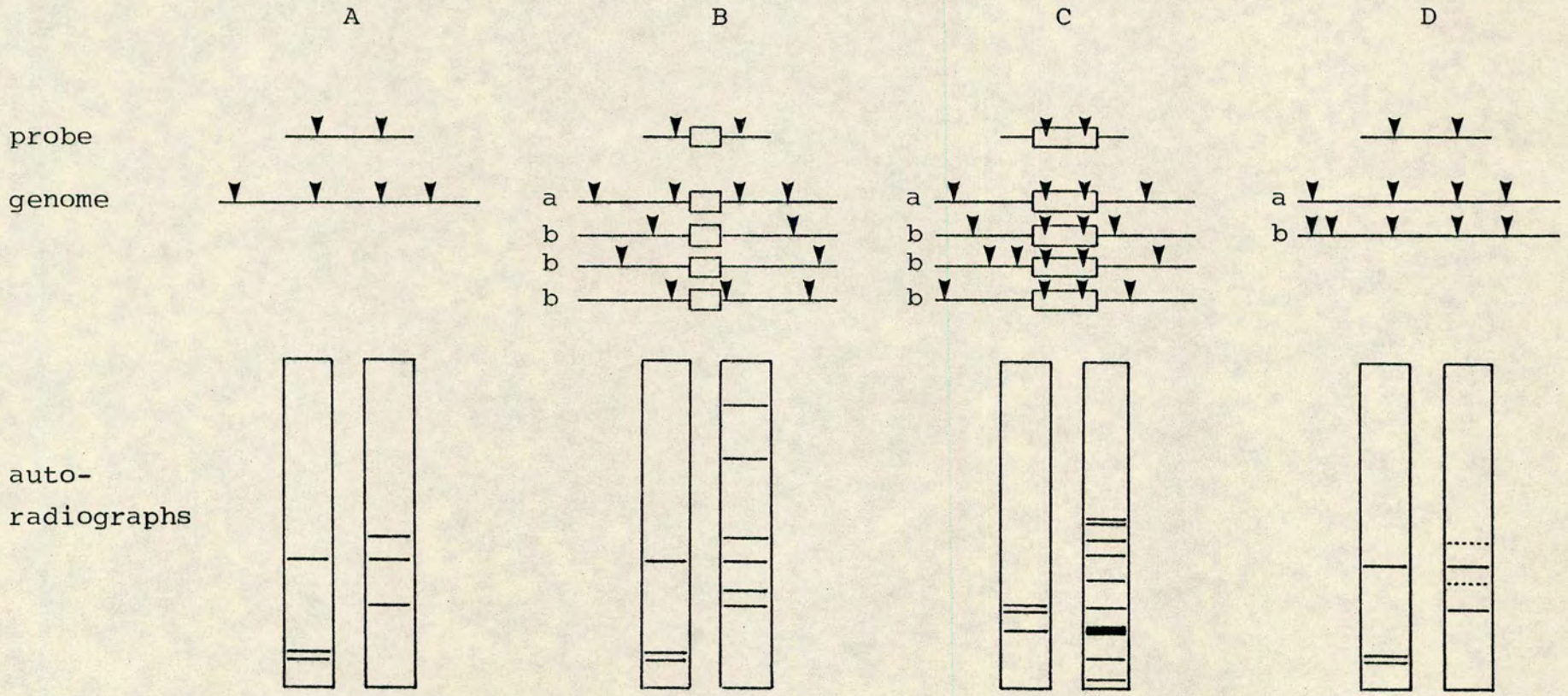
Situation B - The cloned fragment used as probe carries a sequence repeated four times. The repeat does not span a complete restriction fragment. Hence the probe will label six fragments: one internal fragment and two flanking fragments from the sequence from which it was derived (a); and three fragments carrying copies of the repeat from different sites in the genome (b).

Situation C - The cloned fragment used as probe carries a sequence repeated four times. The repeat spans a complete restriction fragment. Hence the probe will label nine fragments: the internal fragment of the repeat from the sequence from which it was derived (a), and from the three copies of the repeat (b), giving a band four times stronger

than expected for single copy; and eight flanking fragments, two from the sequence from which the probe was derived (a), and six from the three copies of the repeat at different sites in the genome (b).

Situation D - The cloned fragment used as probe carries only unique sequence. There is restriction site heterogeneity within the population in the region surrounding this unique sequence: 50% of the population has the pattern of restriction sites shown in (a), and 50% an extra site giving the pattern shown in (b). Hence the probe will label four fragments: one internal fragment, and three flanking fragments, two of the latter hybridising at half the intensity expected for single copy sequences.

Figure 1.2



has a repeated sequence been found at the same site in all individuals of the two Drosophila melanogaster strains Canton S and Oregon R (W. Bender and P. Spierer, personal communication, B. Will and D. Finnegan, personal communication, also this study). Thus by screening a library of a different strain it may be possible to isolate a recombinant carrying sequences from both sides of the repeated region (that is, with the repeat effectively deleted). Fragments from this recombinant can then be used to continue the "walk". It is important to "walk" back to the repeat in the original strain as a control against other sequence rearrangements in the second strain.

The second requirement of a probe is that it does not cross-hybridise with vector or E. coli DNA. Therefore when a recombinant picked up in one screening is to be used as a probe in a subsequent screening, it is necessary to isolate a fragment free of vector DNA. This may be done by either subcloning in a different vector or by purifying a fragment from a gel. It is strongly recommended that the former course be followed. It is difficult to get a completely pure fragment from a gel, and even slight contamination will raise the background level of hybridisation and mask any weak positives. Weak positives are potentially those which will extend the "walk" furthest.

Finally there is the question of whether a sequence is derived entirely from the region in which the "walk" is taking place. There are three circumstances by which a unique sequence from another region of the genome might be present on a probe: use of a fragment from a recombinant isolated with a probe carrying repetitive sequences, use of a fragment from a recombinant carrying a double insert, and use of a fragment from a recombinant carrying an alternative arrangement of sequence to that usually found at this site. The detection of misleading recombinants of this type will be discussed in the section on determining regions of overlap between recombinants.

Screening

The greater the number of plaques screened the more likely it is that a recombinant which will extend the "walk" will be isolated. It is possible to calculate both the probability that a probe will yield a positive or positives, and the expected number of these positives.

Let a be the average length of insert in the recombinant library in kb.
G be the haploid genome size in kb.
N be the number of independent recombinant phage recovered when the library was first manufactured.
n be the number of phage screened.
p be the fraction of the library amplified.
b be the length in kb of the genomic sequence used as probe, given that it is unique.
c be the minimum detectable homology in kb, given that it is less than b.

In order to score as positive a recombinant must carry an insert from a region of the genome which lies across the probe sequence the length of which is given by

$$b + 2(a - c) \qquad (1.6)$$

Inserts which lie within this region will have at least one end which lies within a region of length

$$b + a - 2c \qquad (1.7)$$

Assuming the ends of the inserts in the unamplified library are random (see below), the proportion of inserts in the library which satisfy this condition is given by

$$(b + a - 2c)/G \qquad (1.8)$$

Hence the probability that at least one positive will be detected in the amplified library will be

$$1 - (1 - (b + a - 2c)/G)^{((1 - (1 - 1/Np)^{n})Np)} \qquad (1.9)$$

Compare with equation 1.5

The total number of positives expected with a given probe is

$$n(b + a - 2c)/G \qquad (1.10)$$

The number of different inserts expected among the positives isolated is given by

$$(1 - (1 - 1/Np)^{n})Np(b + a - 2c)/G \qquad (1.11)$$

Two assumptions have been made. The first, that all recombinants are amplified equally, has already been discussed at length in the section dealing with libraries. Amplification and viability effects will reduce the number of different recombinants in the plaques being screened. Differential viability will also exert an effect during the screening process. A poorly growing phage may give a small plaque which is less likely to give a detectably positive signal. This will effectively reduce the number of recombinants being screened. Hence the probability of obtaining any positives, the total number of positives, and, especially, the number of different positives will tend to be slightly lower than predicted by equations 1.9, 1.10, and 1.11.

The second assumption is that the ends of the inserts in the unamplified library are random. If the inserts were generated by shearing of DNA, they should have random ends. However, unless the genome consists of one or more circular molecules, which seems unlikely, there will be specific ends in the DNA prior to shearing: the ends of the chromosomes. Provided that the product of the number of specific ends and the average length of insert is small compared to the haploid genome size, the effect will be negligible, and the equations can be used as given. There is one exception to this, and that is where the sequence being used as probe lies close to a specific end. In such cases the number of recombinants recovered will be lower than otherwise expected.

The sensitivity of the screening process is remarkable. Evidence will be presented that, under the conditions used in this study, overlaps as small as 0.2kb can be detected ($c = 0.2$). Factors affecting the sensitivity are the

viability of the recombinants (see above), the purity of the probe (see above), and, of course, the specific activity of the probe. The necessity of picking any plaque which shows even the slightest hint of hybridisation must be stressed. A weak positive potentially extends the "walk" furthest. Of course, many false positives whose hybridisation is due to an artifact of the screening process will be isolated but most of these will be eliminated during plaque purification.

Determination of regions of overlap.

The most useful way to determine the regions of overlap between recombinants is as follows: DNA from the recombinants to be characterised is digested with a restriction enzyme or enzymes and the fragments electrophoresed through an agarose gel, transferred to nitrocellulose, and probed with nick translated DNA from a recombinant which has already been characterised. Analysis of the fragments which co-migrate and/or cross-hybridise allows areas of overlap and probable directions of extension to be recognised. In this way those recombinants worthy of further investigation can be picked out. Examples of this process are given in section 4.1.

In assessing the validity and value of this technique three factors must be considered: the degree of conservation of restriction sites between strains and individuals of a strain, the possible repetition of sequences within the probes used in the isolation of the recombinants, and the possibility of double inserts.

If the restriction sites were not conserved this method would be almost entirely useless. Extensive restriction mapping and cross-hybridisation would be required before the regions of overlap could be determined and the most useful recombinants identified. Fortunately the restriction sites are very largely conserved (see sections 4.2 and 4.4). Once the recombinants of interest have been identified, the regions of overlap should be confirmed with additional restriction enzymes.

Suppose that the terminal fragment or fragments of one recombinant were to hybridise with the terminal fragment or fragments of another, and that all non-hybridising fragments lie at opposite ends of the common region. This is consistent with their being derived from overlapping regions of the genome. However, it could be argued that either a repeat lay across the common region, or that one of the two recombinants carried a double insert or an uncommon alternative arrangement of sequence, and thus that the non-overlapping regions were derived from parts of the genome which are usually widely separated. This can be countered by the finding of further recombinants which overlap in the manner described above, which would effectively lengthen the region required to be repeated, and limit the non-overlapped region in which the double insert junction could occur. If the non-overlapped region is used as a probe for a subsequent screening of the library, a recombinant stretching back into the overlapped region must be isolated if the linearity of the "walk" is to be proved.

Now suppose it were found that two recombinants, having cross-hybridising fragments, had, adjacent to these fragments, and extending in the same direction, fragments which did not cross-hybridise. If the non-hybridising fragments extend on both sides of the hybridising fragments the presence of a repeat is the most likely explanation, although an alternative arrangement of sequence at this site within the population cannot be excluded completely. If the non-hybridising fragments extend on only one side of the common sequence in one or both recombinants, then several explanations are possible: repetitive sequences, alternative arrangements of sequence, and double inserts. It is not possible to distinguish between these alternatives, and fragments of such recombinants should not be used as probes for subsequent screenings of the library. There is one exception to this, and that is where one switched strains to avoid a long repetitive region (see above).

In this case a recombinant carrying an alternative sequence arrangement is being sought. In Drosophila (and other dipteran flies) it is possible to determine the chromosomal site of origin of a fragment of DNA by in situ hybridisation to the polytene chromosomes found in certain tissues, notably the larval salivary glands. It is recommended that this technique be used to check that any fragment from such a recombinant which is to be used as probe comes from the region of the "walk". Note that the low resolution of in situ hybridisation means that even if the fragment does label the correct site, it is still necessary to "walk" back to the site of the rearrangement in the original strain if the linearity of the "walk" is to be proven.

An alternative way by which the regions of overlap could be determined is by examining heteroduplexes of the recombinants under the electron microscope. This method gives much greater resolution than cross-hybridisation enabling small rearrangements to be detected. The drawback is that in 50% of cases, the inserts will lie in the opposite orientations with respect to the vector sequences, and this will prevent the formation of useful heteroduplexes.

As mentioned earlier, much of our current knowledge of the organisation of the genome comes from the application of recombinant DNA technology. The technique of "walking along the chromosome" extends this in two ways.

(1) The sequence organisation of a large stretch of DNA can be studied, and the inter-relationships (if any) of transcribed sequences, repetitive sequences, and structural features of chromatin may be determined (Yen and Davidson, 1980; W. Bender and P. Spierer, personal communication). Ultimately, one would hope to have a map of the region under study on which, repetitive, transcribed and translated sequences would be identified, sites and regions of nuclease sensitivity located, and the extents of the supercoiled domains (were these to be sequence specific) defined. In Drosophila melanogaster it may also be possible to align such a physical map with the genetic map. Certain tissues, notably the larval salivary glands, contain polytene chromosomes (Daneholt, 1975). These consist of many copies of each interphase chromosome lying side by side and aligned such that normally invisible structural features give rise to a specific, and effectively invariant, pattern of bands. This pattern can be used to identify a particular region of the chromosome (Bridges, 1935; Lefevre, 1976). Many genetic loci have been located to within a few bands of the polytene chromosomes (Lindsley and Grell, 1968). Any cloned fragment of DNA can be located on the polytene chromosome banding pattern by in situ hybridisation (Pardue et al, 1973; Wensink et al, 1974). Hence it follows that rough alignment of the genetic and physical maps is possible. This will be particularly useful where a region of the genome has been "saturated" with mutations (Judd et al, 1972).

(2) Genes which have only been identified genetically, and about which nothing is known biochemically, may be cloned. This aspect of the technique hinges upon the alignment of the genetic and physical maps. Two examples are known: the cloning of bithorax (Lindsley and Grell,

1968) from Drosophila melanogaster, by W. Bender and P. Spierer (personal communication) where alignment was achieved by in situ hybridisation to the polytene chromosomes as described above; the cloning of the yeast centromere (Clarke and Carbon, 1980) where alignment was achieved through knowledge of the biochemical functions of the surrounding loci.

In the absence of this approach, only those genes for which a specific probe can be prepared, or which can be expressed and selected in E. coli or yeast may be cloned. The former class includes genes which are transcribed to give a major proportion of the RNA of some cell type (for example, globin, ovalbumin, histone, heat shock, rRNA and 5S genes) and are therefore unlikely to be typical of the majority of genes. Genes in the latter class will be limited in number, and largely confined to those coding for functions common to higher and lower organisms.

The following sections describe a "walk" in the 84 region of Drosophila melanogaster; part of the right arm of the third chromosome defined according to the polytene banding pattern (Bridges, 1935; Lefevre, 1976). This "walk"is part of a long term project to clone the homeotic locus (homeotic loci are involved in development and differentiation; see Morata and Lawrence (1977)) Antennapedia (Lindsley and Grell, 1968), located at 84 B1 - 2 (Denell, 1973; Duncan and Kaufman, 1975). The starting points for the "walk" were the recombinants pDm2 and 14C4, originally believed to be derived from 84 D. The latter has now been shown to originate from 84 F (see below).

pDm2 (Wensink et al, 1974)

This consists of an 8.6kb insert of Drosophila melanogaster embryo DNA from strain Oregon R cloned by oligo dA:dT tailing into the Eco R1 site of pSC101. In situ hybridisation locates the insert sequence at 84 D1 - 2 (Wensink et al, 1974; A. Bowcock and

R. Hodgetts, personal communication).

14C4 (Gehring, 1978; Dudler et al, 1980)

This consists of a 15.85kb insert of Drosophila melanogaster embryo DNA from strain Oregon R cloned by oligo dA:dT tailing into the Eco R1 site of RSF2124. It was isolated in a screening experiment using labelled tRNA, and has been shown to carry genes for asparagine and arginine tRNAs (Dudler et al, 1980). In situ hybridisation originally suggested the insert was derived from 84 D (Gehring, 1978); it is now believed to be from 84 F1 - 2 (Dudler et al, 1980; A. Bowcock and R. Hodgetts, personal communication).

Section 2.

MATERIALS.

2.1 Chemicals.

Chemicals were purchased from the following companies :-

Agarose - Miles Laboratories Ltd.
Ampicillin - Beecham Research Laboratories.
Tetracycline - Lederle Laboratories.
Radioactive compounds - The Radiochemical Centre, Amersham.
"Repelcote" - Hopkin and Williams.

All other chemicals from B.D.H. Chemicals Ltd., Fisons Scientific Apparatus, Koch-Light Laboratories Ltd., and Sigma. Chemicals were of analytical or laboratory reagent grade as appropriate.

2.2 Enzymes.

Enzymes were obtained from the following sources :-

Prepared in this laboratory - Alu 1 (A. Newman); Bam Hl, Eco Rl, Hind lll, Sal 1 (K. Mileham); DNA polymerase 1 (B. Will); Eco Rl (J. Gould); Pst 1 (J. de Banzie); Xho 1 (B. Sain).

Purchased - Bam Hl, Hha 1, Hpa 1, Kpn 1, Sac 1, Sal 1, Sma 1 (New England Biolabs); DNase 1 (Miles Laboratories Ltd.); DNase 1, Lysozyme, Protease (type VI), RNase A (Sigma); Eco Rl, Kpn 1 (Bethesda Research Laboratories); RNase A, Sma 1 (Boehringer Mannheim); Xma 1 (Worthington Diagnostics).

2.3 Other Materials.

Bleach - A.J. Beveridge Ltd.

Developers - Microphen (Ilford Ltd.)
Polycon (May and Baker Ltd.)

Intensifying screens - Cronex Lightning Plus (Du Pont Ltd.)
Ilford Fast Tungstate (Ilford Ltd.)

Nitrocellulose - Millipore UK Ltd.
Schleicher and Schüll GmbH.

Photographic film - FP4 Professional (Ilford Ltd.).

Sephadex G50 - Pharmacia Fine Chemicals.

X-ray film - Fuji RX Medical X-ray Film (Fuji Photo Film Co. Ltd.).
Kodak X-Omat H Film (Kodak Ltd.).

2.4 Media.

BBL agar - Baltimore Biological Laboratories trypticase, 10g; NaCl, 5g; Difco agar, 10g per litre.

BBL top agar - As for BBL agar, but only 6.5g Difco agar per litre.

L agar - Difco Bacto Tryptone, 10g; Difco Bacto yeast extract, 5g; NaCl, 10g; Difco agar, 15g per litre, pH 7.2.

L broth - Difco Bacto Tryptone, 10g; Difco Bacto yeast extract, 5g; NaCl, 5g per litre, pH 7.2.

Phage buffer - KH_2PO_4, 3g; Na_2HPO_4 (anhydrous), 7g; NaCl, 5g; 10ml 0.1M $MgSO_4$; 10ml 0.01M $CaCl_2$; 1ml 1% gelatin solution per litre.

Antibiotics added to L agar and L broth as required - Ampicillin, 40mg per litre; Chloramphenicol, 100mg per litre; Tetracycline, 20mg per litre.

2.5 Bacterial Strains.

E. coli strains C600 (Appleyard, 1954) and ED8654 (NM259) (Murray et al, 1977) were used interchangeably for growth of phage on plates and in liquid cultures. Plasmids were maintained in *E. coli* HB101 (Boyer and Roulland-Dussoix, 1969). The genotypes of these strains are given below.

C600 - *leu*$^-$, *sup* E, *thi*$^-$, *thr*$^-$, *ton* A.

ED8654 - *gal*$^-$, *hsd*$_K$ M$^+$, *hsd*$_K$ R$^-$, *met*$^-$, *sup* E, *sup* F, *trp* R$^-$.

HB101 - *ara*$^-$, *arg*$^-$, *gal*$^-$, *hsd*$_B$ M$^-$, *hsd*$_B$ R$^-$, *lac*$^-$, *pro*$^-$, *rec* A$^-$, *Str*^{r}.

2.6 Lambda-*Drosophila* Recombinant DNA Libraries.

The CS library was obtained from T. Maniatis. It consists of sheared fragments of embryo DNA from *Drosophila melanogaster* strain Canton S cloned into the lambda replacement vector Charon 4 (Blattner et al, 1977) using Eco R1 linkers (Maniatis et al, 1978).

The OR library was obtained from the laboratory of D. Hogness. It consists of sheared fragments of embryo DNA from *Drosophila melanogaster* strain Oregon R cloned into the lambda replacement vector Sep 6 using oligo dC:dG tailing (Wensink et al, 1974 ; M. Wolfner, personal communication).

2.7 Buffers.

The following buffers are used in many of the methods described in section 3.

1 x SSC - 0.15M NaCl; 0.015M *tri*-sodium citrate. Prepared as 20 x SSC and diluted as appropriate.

TE - 10mM Tris-HCl; 1mM EDTA: pH 8.0.

All other solutions are described in the relevant methods.

Section 3.

METHODS.

3.1 Plating Cells.

A fresh overnight culture of E. coli C600 or ED8654 was diluted 25 fold into L broth. After 2 hours of aerated growth at 37°C, the cells were pelleted and resuspended in 1mM $MgSO_4$. Cells prepared in this manner were stored at 4°C and could be used for up to 10 days thereafter.

3.2 Phage Titration.

A sample of the phage stock to be titred was diluted as appropriate with phage buffer and 0.2ml of this dilution mixed with 0.2ml plating cells. After allowing to stand for 20 minutes to permit adsorption of the phage to the bacteria, 2.5ml of molten BBL top agar (45°C) was added and the mixture poured onto a BBL agar plate. Plates were incubated overnight at 37°C and the plaques counted.

3.3 Plaque Hybridisation.

The method used was a modification of that of Benton and Davis (1977). Phage were grown on plating cells as described in section 3.2, but using 0.7% agarose in 10mM $MgCl_2$ in place of BBL top agar, and dry BBL agar plates. After the overnight incubation, a disc of nitrocellulose was placed on each plate and two holes punched through it into the plate with a syringe needle to facilitate orientation later. After 2 minutes the disc was carefully removed and placed plaque side up on a pad of blotting paper soaked in 0.5M NaOH; 1.5M NaCl for 2 minutes. The filter was then transferred to a solution of 0.5M Tris-HCl; 3M NaCl; pH 7.0. When filters had been prepared from all plates they were rinsed in 2 x SSC, blotted dry, and baked in a vacuum oven at 80°C for 1½ hours. Filters were hybridised and autoradiographed

as detailed in sections 3.17 and 3.18.

The use of 0.7% agarose in 10mM $MgCl_2$ instead of BBL top agar minimised the tendency of the top layer to adhere to the nitrocellulose filter and peel off the plate when the filter was being removed.

3.4 Plate Lysates.

Phage lysates were prepared by mixing 10^6 PFU phage with 0.2ml plating cells, allowing the phage to adsorb for 20 minutes, adding 2.5ml molten BBL top agar, and pouring the mixture onto a fresh L agar plate. Plates were then incubated at 37°C. Phage were harvested by either (a) pipetting 3 to 5ml L broth onto the plates when the plaques became confluent (after about 7 hours), storing overnight at 4°C, then pipetting off the liquid, or (b) incubating the plates overnight at 37°C then scraping off the top layer with a sterile pipette, rinsing the plate with 3 to 5ml L broth, then pelleting the top layer through the L broth, decanting and retaining the supernatant. The lysate was stored at 4°C with a few drops of chloroform, whichever method was used.

3.5 Preparation of Phage DNA from Plate Lysates.

Plating cells were infected with 10^6 PFU phage as in section 3.4, but using L agarose top layer (0.65% agarose in L broth) in place of BBL top agar, and fresh L agarose plates (1.5% agarose in L broth) in place of L agar plates. When confluent lysis was achieved (after about 7 hours incubation at 37°C) the plates were over-layered with 3 to 5ml of 10mM Tris-HCl; 10mM $MgSO_4$; pH 7.5, or 10mM Tris-HCl; 10mM EDTA; pH 7.5 (the former buffer allows excess lysate to be used as a normal phage lysate, the latter does not) and stored at 4°C overnight. The liquid was then drawn off each plate, and to a 1ml aliquot was added 0.1ml 0.5M EDTA (pH 8.5), 0.05ml 2M Tris-base, 0.05ml 10% SDS, and 2.5ul diethylpyrocarbonate. After mixing, the samples were heated at 65°C for 30 minutes in open tubes, chilled on

ice, and 0.25ml 5M CH_3COOK added. The samples were allowed to stand on ice for 1 hour then centrifuged. The supernatant was decanted and DNA precipitated with 2 volumes of ethanol. The precipitate was re-dissolved in a small volume of TE buffer.

This method was used as a rapid means of preparing small quantities of DNA from a large number of phage isolates for characterisation before making large scale preparations. However the method was found to give very erratic results with respect to both quantity and quality of phage DNA obtained.

3.6 Preparation of Phage DNA.

A fresh overnight culture of E. coli C600 or ED8654 was diluted 20 fold into 200ml L broth, supplemented with 10mM $MgCl_2$, and grown at 37°C with aeration until an A_{650} of between 0.45 and 0.6 was reached. The culture was then infected with 4 to 6 x 10^6 PFU of phage and incubated at 37°C with aeration until lysis occurred (2½ to 3 hours). The culture was shaken for a further 10 minutes with 0.5ml chloroform. After addition of 10g NaCl and 200ug of DNase 1 and RNase A, the culture was allowed to stand at room temperature for 1 hour before the cell debris was pelleted. Polyethylene glycol 6000 was then added to 10% (w/v) concentration and the phage suspension stored at 4°C for 15 to 20 hours. The precipitate was pelleted and resuspended in 5ml phage buffer at 4°C, then loaded onto a step gradient of CsCl in phage buffer with steps of 1.3g per ml, 1.5g per ml, and 1.7g per ml. After centrifugation at 33krpm for 2 hours at 20°C in an MSE 6 x 14 swing-out rotor the phage band was removed by syringe through the wall of the tube and dialysed against TE buffer for at least 1 hour. The purified phage were then extracted 3 or 4 times with redistilled phenol equilibrated with TE buffer, the phenol back-extracted with TE buffer, and the combined aqueous phases dialysed extensively against TE buffer.

3.7 Preparation of Plasmid DNA.

All media were supplemented with antibiotics as required. An overnight culture of the plasmid carrying strain was diluted 50 fold into 1 litre of L broth and grown at 37°C with aeration until an A_{650} of 1 was reached (3 to 4 hours). Chloramphenicol (100mg) was added and the culture grown for a further 12 hours. The cells were pelleted, resuspended in 10ml sucrose mix (25% sucrose; 50mM Tris-HCl; 10mM EDTA; pH 8.1), and 3ml of 10mg per ml lysozyme in sucrose mix added. After 5 minutes on ice, 3ml of 0.5M EDTA (pH 8.1) was added, and after a further 5 minutes on ice, 27ml Triton mix (0.1% Triton X-100; 50mM Tris-HCl; 62.5mM EDTA; pH 8.1). The solution was allowed to stand on ice for 10 minutes and then centrifuged at 25krpm for 30 minutes at 4°C in an MSE 6 x 14 swing-out rotor. The supernatant was decanted and, after addition of 0.95g CsCl and 0.1ml 10mg per ml ethidium bromide per ml, was centrifuged at 38krpm for 72 hours at 20°C in an MSE 8 x 40 angle rotor. The gradient thus generated was viewed under long wavelength UV light, and the supercoiled plasmid DNA band removed by syringe through the side of the tube. Ethidium bromide was removed from the DNA on a small column of 'Dowex' 50W-X8 and the DNA dialysed overnight against TE buffer before being precipitated with ethanol and re-dissolved in TE buffer.

3.8 Preparation of Drosophila melanogaster Embryo DNA.

Flies were grown at 25°C on a medium consisting of 6.43g agar; 21.43g dried yeast; 107.14g sugar; 118.54g cornmeal; 30ml 10% Nipagin in 95% ethanol per litre of water. Eggs were collected over a period of 14 to 15 hours on 1% agarose plates spread with yeast paste, then the plates removed and stored at room temperature for 2 to 3 hours. The eggs were washed through a coarse filter onto a fine filter, rinsed with distilled water, floated on 25% sucrose solution to remove yeast granules,

collected on a fine filter, and washed with 0.7% NaCl; 0.01% Triton X-100. The embryos were dechorionated in a 50% bleach solution for 3 minutes at room temperature, collected on a fine filter, and washed thoroughly with 0.7% NaCl; 0.01% Triton X-100. Embryos prepared in this manner were stored at -20°C for up to six months before use.

Embryos were thawed in buffer A (0.25M sucrose; 30mM Tris-HCl (pH 7.5); 10mM EDTA; 2.5mM $CaCl_2$), disrupted in a Dounce homogeniser, and filtered through a fine filter. Nuclei were pelleted by centrifugation at 4 krpm for 15 minutes at 4°C in a Sorvall HB-4 4 x 50 swing-out rotor. The pellet was washed twice with buffer A, resuspended in 0.15M NaCl; 50mM Tris-HCl (pH 8.0); 0.1M EDTA, and SDS added to 0.5%. The suspension was incubated at 60°C for 10 minutes, cooled, and RNase A (heat treated at 80°C for 10 minutes) added to 0.25mg per ml. After incubation at 37°C for 45 minutes, protease (autodigested for 2 hours at 37°C before use) was added to give 2mg per ml, and the mixture returned to 37°C for a further 15 to 20 hours. The mixture was then extracted with re-distilled phenol equilibrated with TE buffer until the interface was free of protein, the phenol layers back-extracted with TE buffer, and the aqueous phases pooled and dialysed against several changes of TE buffer. The DNA was finally precipitated with ethanol and re-dissolved in TE buffer.

3.9 Ethanol Precipitation of DNA.

The salt concentration of the DNA solution was raised by addition of one tenth volume of 2.7M sodium acetate; 0.1M magnesium acetate; pH 5.0 if necessary. Two volumes of ethanol were added and the DNA allowed to precipitate at -20°C for at least 2 hours. The DNA was then pelleted in either plastic or siliconised glass centrifuge tubes, dried under vacuum, and re-dissolved in TE buffer.

3.10 Restriction Digestion.

The following conditions were used for digestion of DNA with restriction enzymes :-

Alu 1 - 50mM NaCl; 6mM Tris-Hcl (pH 7.5); 6mM $MgCl_2$; 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Bam Hl - as Alu 1.

Eco Rl - 100mM Tris-HCl (pH 7.5); 10mM $MgCl_2$; 100ug per ml gelatin. 37°C.

Hha 1 - as Alu 1.

Hind 111 - 60mM NaCl; 10mM Tris-HCl (pH 7.5); 6mM $MgCl_2$; 100ug per ml gelatin. 37°C.

Hpa 1 - 20mM KCl; 10mM Tris-HCl (pH 7.5); 10mM $MgCl_2$; 1mM dithiothreitol; 100ug per ml gelatin. 37°C.

Kpn 1 - 6mM NaCl; 6mM Tris-HCl (pH 7.5); 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Pst 1 - as Alu 1.

Sac 1 - as Hind 111.

Sal 1 - 100mM NaCl; 6mM Tris-HCl (pH 8.0); 6mM $MgCl_2$; 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Sma 1 (Boehringer-Mannheim) - 15mM KCl; 15mM Tris-HCl (pH 8.5); 6mM $MgCl_2$; 100ug per ml gelatin. 25°C.

Sma 1 (New England Biolabs) - 20mM KCl; 6mM Tris-HCl (pH 8.0); 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Xho 1 - 150mM NaCl; 6mM Tris-HCl (pH 8.0); 6mM $MgCl_2$ 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Xma 1 - 6mM Tris-HCl (pH 8.0); 6mM $MgCl_2$; 6mM 2-mercaptoethanol; 100ug per ml gelatin. 37°C.

Double and triple digestions were carried out simultaneously or sequentially depending on the assay

conditions of the enzymes involed. Reaction volumes were from 0.01 to 0.1ml, incubation periods from 1 to 4 hours. Reactions were terminated by heating at 70°C for 5 minutes.

3.11 Agarose Gel Electrophoresis.

Samples of DNA were prepared for electrophoresis by addition of one third volume of tracking dye (0.1% bromophenol blue; 10mM EDTA; 20% glycerol) and, in the case of phage DNA, heating to 70°C for 5 minutes immediately before loading onto the gel. This latter treatment is necessary to prevent the cohesive ends of lambda from annealing.

Electrophoresis was through horizontal agarose slabs (21.5cm x 13cm x 0.5cm) using either Tris-acetate (40mM Tris; 20mM CH_3COONa; 1mM EDTA; pH 8.2) or Tris-borate (90mM Tris; 90mM boric acid; 2.75mM EDTA) buffer. Tris-borate buffer is prone to give slight smearing if the sample wells are not kept full while the samples are being run into the gel; otherwise the results are indistinguishable. Samples were run in at 3 volts per cm and the voltage then turned down to 1 to 1.5 volts per cm to give a running time of 14 to 20 hours as desired. Electrophoresis was stopped when the tracking dye had reached the end of the gel.

Visualisation of the DNA was by incorporation of ethidium bromide (50ug) into the gel itself or by staining after electrophoresis for 30 minutes in a 2mg per litre ethidium bromide solution followed by 30 minutes in distilled water. The latter technique is better when rapidly migrating fragments are present.

Using Tris-borate buffer, a few gels were run in 2 to 3 hours at 15 volts per cm, bathed in buffer to prevent over-heating. This method is quite satisfactory for gels of 1% agarose and above; below this concentration skewing of the tracks becomes a problem.

Digests of lambda $c_1$857 S7 DNA with Hind 111 and/or Eco R1, and of pBR322 DNA with Alu 1 were used as

standards for molecular weight determination.

3.12 Polyacrylamide Gel Electophoresis.

Samples were prepared as for agarose gel electrophoresis. Vertical polyacrylamide slab gels (15.5cm x 15cm x 0.15cm) were prepared by polymerisation of 50ml acrylamide solution (prepared by dilution of a 20% stock 20:1 acrylamide:bisacrylamide solution with Tris-borate buffer (see section 3.11)) by addition of 0.03ml TEMED and 0.15ml 10% ammonium persulphate solution. The running buffer was also Tris-borate. Samples were run in at 5 volts per cm and then electrophoresed at 2.5 volts per cm until the tracking dye had reached the bottom of the gel (14 to 18 hours).

DNA fragments were visualised by staining after electrophoresis for 30 minutes in a 2mg per litre solution of ethidium bromide followed by 30 minutes in distilled water.

Molecular weight standards used were lambda $c_1$857 S7 DNA digested with Hind 111 and Eco R1, and Col E1 DNA digested with Hha 1.

3.13 Gel Photography.

Gels were photographed under short wavelength UV light onto FP4 Professional 5" x 4" sheet film using a Rodenstock Ysarex 150mm f/4.5 lens fitted with a Hoya R(25) red filter. Exposure was 10 minutes at f/4.5. Development was for 10 minutes in Microphen.

Relative mobilities of the fragments were measured directly from the negative.

3.14 Extraction of DNA from Agarose Gels.

For isolation of a particular fragment of a recombinant, 100ug of DNA was digested with the appropriate restriction enzyme and loaded onto an agarose gel. The agarose gel was prepared, run and photographed in the normal manner (see sections 3.11 and 3.13). The gel was then viewed under long wavelength UV light and the strip of gel

containing the fragment desired cut out using a sterilised razor blade. The gel strip was cut into 1mm thick slices and placed in dialysis tubing containing 2ml electro-elution buffer (5mM Tris-base; 2.5mM CH_3COOH) and the dialysis tubing sealed and submerged in electro-elution buffer in a 30cm long tank with electrodes at either end. A potential of 200 volts was applied across the tank, the dialysis bag being fixed at right angles to the flow of current. After two hours the polarity was reversed for 5 to 10 minutes, and then reversed again for a further 5 to 10 minutes. The liquid in the bag was now removed by syringe and the bag rinsed with a further 1ml electro-elution buffer. The combined solutions were extracted with re-distilled phenol equilibrated with TE buffer, and dialysed against a suspension of 'Dowex' 50W-X8 in 0.5M NaCl followed by several changes of TE buffer. Finally the DNA was ethanol precipitated and re-dissolved in a small volume of TE buffer.

DNA prepared in this manner was a poor but adequate substrate for nick translation, giving incorporations of 3 to 8%. Yield of purified fragment was between 25 and 40%.

3.15 Nick Translation.

A modified version of the method of Rigby et al (1977) was used. Between 0.25 and 1ug of DNA was nick translated in a 21ul reaction mix containing 50mM Tris-HCl; 5mM $MgCl_2$; 1.5mM 2-mercaptoethanol; 0.015mM each of unlabelled dATP, dCTP, dGTP, and dTTP; (α^{32}P)dCTP or dGTP; 5ug per ml bovine serum albumin; pH 7.2. The reaction was started by addition of 20pg DNase 1 and 0.5 units of DNA polymerase 1, and incubated at 15°C. The amount of labelled nucleotide used was between 5 and 50uCi (specific activity 350 to 400Ci per mmol) depending on the specific activity required of the probe. Where a high specific activity was essential, for example for

probes to transfers of total Drosophila DNA, the unlabelled nucleotide corresponding to the labelled nucleotide was omitted from the reaction mix.

After incubation for 1 hour, a sample was taken and incorporation of ^{32}P into acid precipitable material determined. If this was sufficient for the experiment in hand, the reaction was terminated by addition of 10ug of sonicated calf thymus DNA in 200ul TE buffer and immediate extraction with re-distilled phenol equilibrated with TE buffer. The phenol layer was back-extracted with TE buffer, the combined aqueous phases passed over a Sephadex G50 column (10 x 1cm) and the first peak of labelled material collected. If the incorporation was not satisfactory after the 1 hour assay, the incubation was continued for up to 4 hours, with addition of further DNase 1 and/or DNA polymerase 1 as necessary.

Incorporations of up to 50% and specific activities of up to 4×10^7 cpm per ug DNA were obtained by this method.

3.16 Transfer of DNA from Agarose Gels to Nitrocellulose Filters.

The method used was essentially that of Southern (1975). Agarose gels were prepared, run, and photographed as described in sections 3.11 and 3.13, then soaked in 0.5M NaOH; 1.5M NaCl for 45 minutes followed by 45 minutes in 0.5M Tris-HCl; 3M NaCl; pH 7.0. The gel was then placed on a wick of blotting paper soaked in 20 x SSC and dipping into a reservoir of 20 x SSC. A piece of nitrocellulose, wetted in 2 x SSC, was placed on top of the gel, care being taken to exclude all air bubbles, and trimmed to the size of the gel. A stack of blotting paper, 2 to 3cm thick, was placed on top of the nitrocellulose and the assembly left to transfer. After 16 to 20 hours the nitrocellulose filter was removed, rinsed in 2 x SSC, blotted dry, and baked at 80°C in a vacuum oven for 1½ hours. The filter could then be stored at 4°C until required.

3.17 Hybridisation of Labelled Nucleic Acid to DNA on Nitrocellulose Filters.

Filters for hybridisation, prepared as described in sections 3.3 and 3.16, were first washed in 50% formamide; 4 x SSC; 1 x Denhardt mix (0.02% bovine serum albumin; 0.02% Ficoll; 0.02% polyvinyl pyrrolidone (Denhardt, 1966)) at 37°C for between 2 and 20 hours. After transfer to fresh buffer, probe DNA containing sufficient sonicated calf thymus DNA to give a final concentration of 15ug per ml, the mixture having been denatured by heating at 95°C for 10 minutes, was added. When labelled RNA was used as probe, Drosophila melanogaster embryo rRNA was used in place of sonicated calf thymus DNA, and the mixture was not heat treated. After incubation at 37°C for 16 to 40 hours, the filters were washed twice for 30 minutes in fresh 50% formamide; 4 x SSC, with or without Denhardt mix (no difference could be detected) at 37°C, and then for 2 to 3 hours in 2 x SSC at room temperature with 4 or 5 changes of solution. Having been blotted dry the filters were ready for autoradiography.

3.18 Autoradiography.

Hybridised filters were autoradiographed with or without an intensifying screen as required. If an intensifying screen was used the film was pre-fogged and exposed at -70°C (Laskey and Mills, 1977). Exposures varied from 15 to 150 hours. Films were developed for 10 minutes in Polycon developer.

3.19 Electron Microscopy.

DNA was spread for electron microscopy according to the method of Davis et al (1971). The contour lengths of ten molecules were measured and the molecular weights determined by comparison with standards spread on the same grid. Molecular weight standards used were pSC101 DNA (9.22kb, Wensink et al, 1974) and X174 DNA (5.375kb, Sanger et al, 1977).

Section 4.

RESULTS.

4.1 Isolation of Recombinants.

This section describes the three screenings of the CS and OR libraries which were carried out, and the determination of the regions of overlap of the recombinants which were isolated. The detailed characterisation of the most interesting of these recombinants, including the location of repetitive sequences, is set out in the sections which follow. However, since detailed analysis of a recombinant is required before it (or a fragment of it) can be used as a probe in a screening of a library, this is obviously not the order in which the experiments took place. For example, pDm2 and 14C4 were examined for repetitive sequences before being used in the first screening. Although this order of presentation may necessitate a certain amount of referral to later sections (the summary in section 4.8 may be found useful in this respect), it is felt that this disadvantage will be outweighed by the coherence given to the later sections.

First screening experiment.

In the first screening, 50 000 plaques of the CS library were probed with a mixture of nick translated pDm2, 14C4, crDm655, and crDm656 DNAs. The latter are recombinants which carry fragments of cDm412 (fragments F and C respectively, see Rubin et al, 1976, Finnegan et al, 1977). Of the 179 plaques which initially gave positive hybridisation, 121 hybridised with a mixed probe of crDm655 and crDm656 DNA, and 11 with a mixed probe of pDm2 and 14C4 DNA after plaque purification (see table 4.1). The pDm2 and 14C4 homologous recombinants were designated CS001 to CS011. None of them showed hybridisation with crDm655 or crDm656 DNA.

DNA was prepared from plate lysates of CS001 to CS011 and digested with Eco R1, electrophoresed through 0.7% agarose gels, and the fragments transferred to nitrocellulose and probed with nick translated pDm2 or 14C4 DNA. From the results obtained (see figures 4.1 and 4.2) it is clear that CS001, CS002, CS004, CS005, CS007, CS008, CS009, and CS011 have homology only with the pDm2 probe, and CS003, CS006, and CS010 only with the 14C4 probe. The 19 and 11kb fragments present in all tracks are vector DNA.

I will now give the reasoning by which the positions and extents of the overlapping regions of these recombinants were deduced. The conclusions reached are presented diagramatically in figures 4.3 and 4.4; it may be found helpful to consult these while following the arguments set out below. The orders of the Eco R1 fragments in pDm2 and 14C4 had been determined previously (see section 4.2, figures 4.19 and 4.20).

Considering the pDm2 homologous recombinants first, the 2.05, 1.15, and 0.55kb fragments which hybridise with pDm2 DNA are taken to be equivalent to the 2.05, 1.15, and 0.55kb fragments of pDm2. CS001, CS002, CS004, CS008, and CS011 carry these fragments plus two others. In all cases the larger fragment shows very little hybridisation with

Table 4.1

Details of the first screening experiment.

Library	CS			
Probes	pDm2	14C4	crDm655	crDm656
Number of plaques screened	(	50 000		)
Probability of positive(s)[a]	100%	100%	100%[d]	100%[d]
Total positives expected[b]	7	10	(about 200[d]	)
Different positives expected [c]	7	9	(?	)
Initial number of positives	(	179		)
Final number of positives	8	3	(121	)
Number of different positives	6	2	(?	)

Notes

(a) Calculated using equation 1.9 with the following values:-

a = 16 (Maniatis et al, 1978)
G = 165 000 (Rudkin, 1972)
N = 600 000 (Maniatis et al, 1978)
n = 50 000
p = 1
b = 8.6 for pDm2, 15.85 for 14C4
c = 0.2

(b) Calculated using equation 1.10 with the same values as in (a) above.

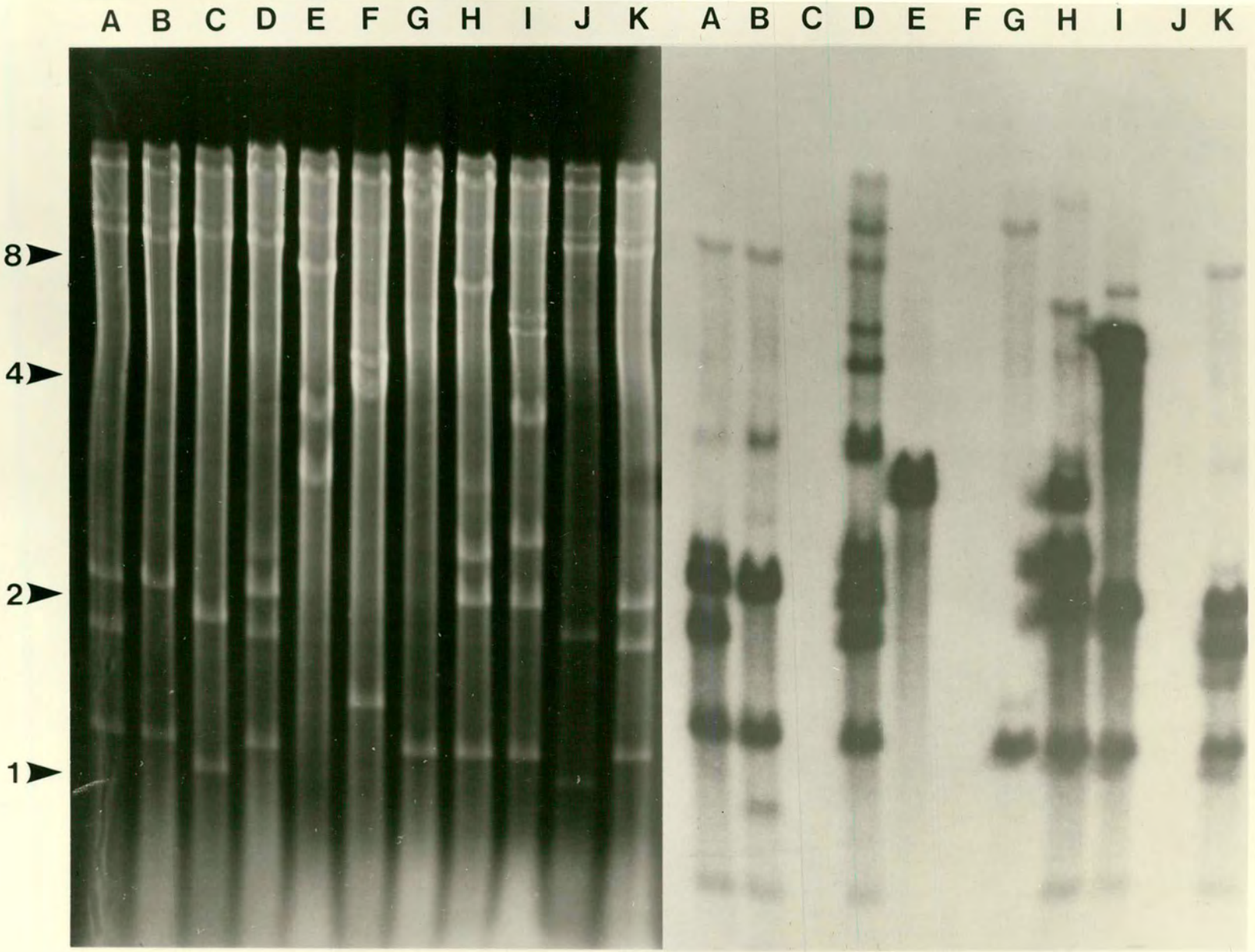
(c) Calculated using equation 1.11 with the same values as in (a) above.

(d) These figures are rough estimates. The repetition frequency of the cDm412 element was taken to be about 40.

Figure 4.1

Plate lysate DNA from recombinants CS001 to CS011 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated pDm2 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate approximate fragment sizes in kb.

Track	Recombinant
A	CS001
B	CS002
C	CS003
D	CS004
E	CS005
F	CS006
G	CS007
H	CS008
I	CS009
J	CS010
K	CS011

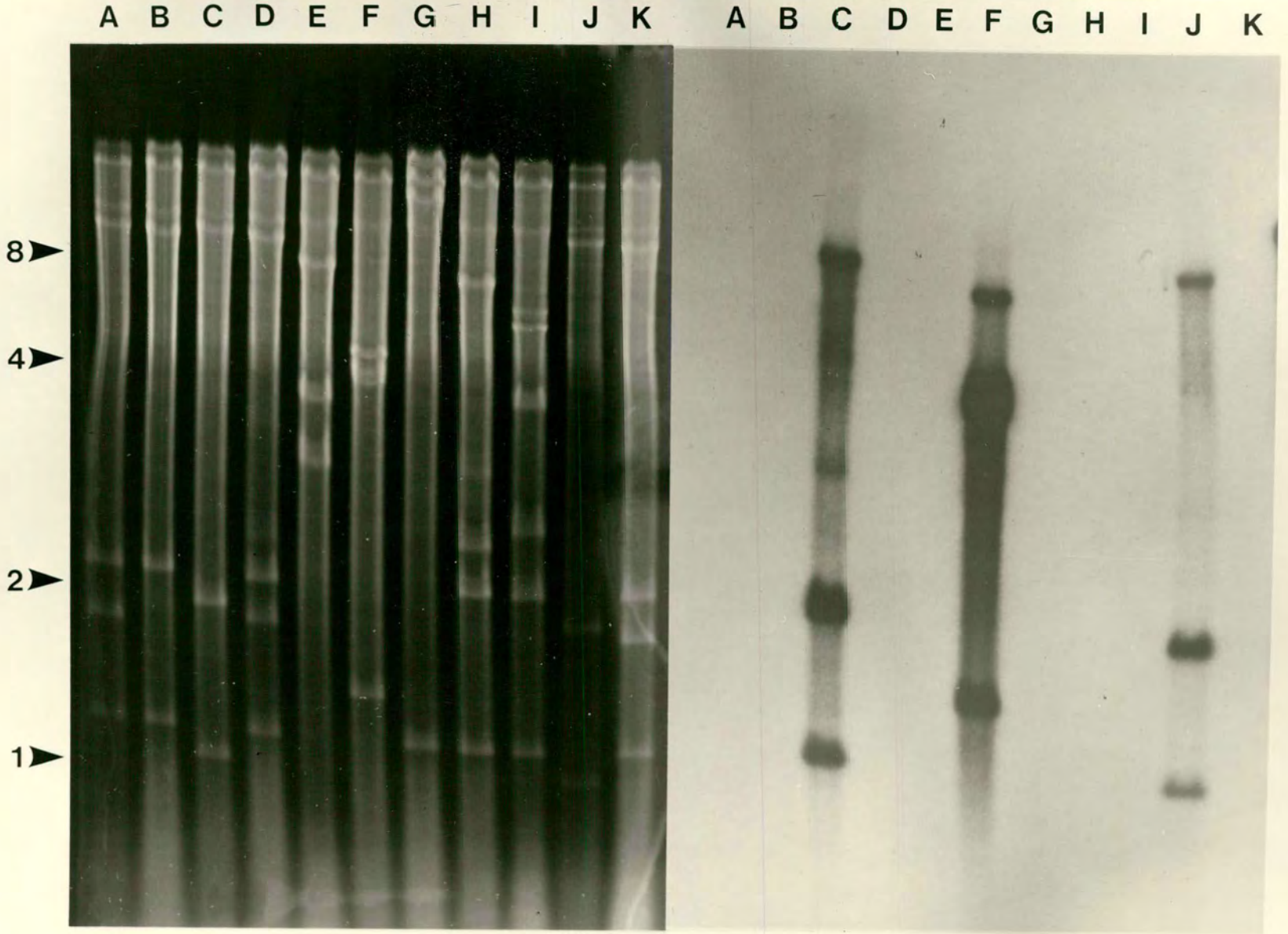


4·1

Figure 4.2

Plate lysate DNA from recombinants CS001 to CS011 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated 14C4 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate approximate fragment sizes in kb.

Track	Recombinant
A	CS001
B	CS002
C	CS003
D	CS004
E	CS005
F	CS006
G	CS007
H	CS008
I	CS009
J	CS010
K	CS011



4·2

pDm2 DNA, and the smaller extensive hybridisation. Hence the larger fragment must extend beyond the end of pDm2, to the left (as the map is drawn) of the 1.15kb fragment, and the smaller lie within pDm2 to the right of the 0.55kb fragment. CS001, CS004, and CS011 share the same additional two fragments; they carry the same insert. The smaller fragment of CS002 is not visible on the gel photograph, but a band of about 0.8kb is present on the autoradiograph. This band must presumably be lost in the background of E. coli DNA in the photograph.

The CS005 track has no fragments which co-migrate with those of pDm2, and the only fragment with homology is about 3kb long. Hence this fragment must extend either to the left or right of pDm2. As there are at least two Eco R1 sites within the insert of CS005, and since, from the recombinants CS001, CS004, and CS011, it is at least 9kb to the nearest Eco R1 site left of pDm2, CS005 must extend beyond the right end of pDm2. This conclusion is supported by the level of hybridisation to the 3kb fragment, which is higher than that found for the fragments of CS001, CS002, CS004, CS008, and CS011 which extend to the left.

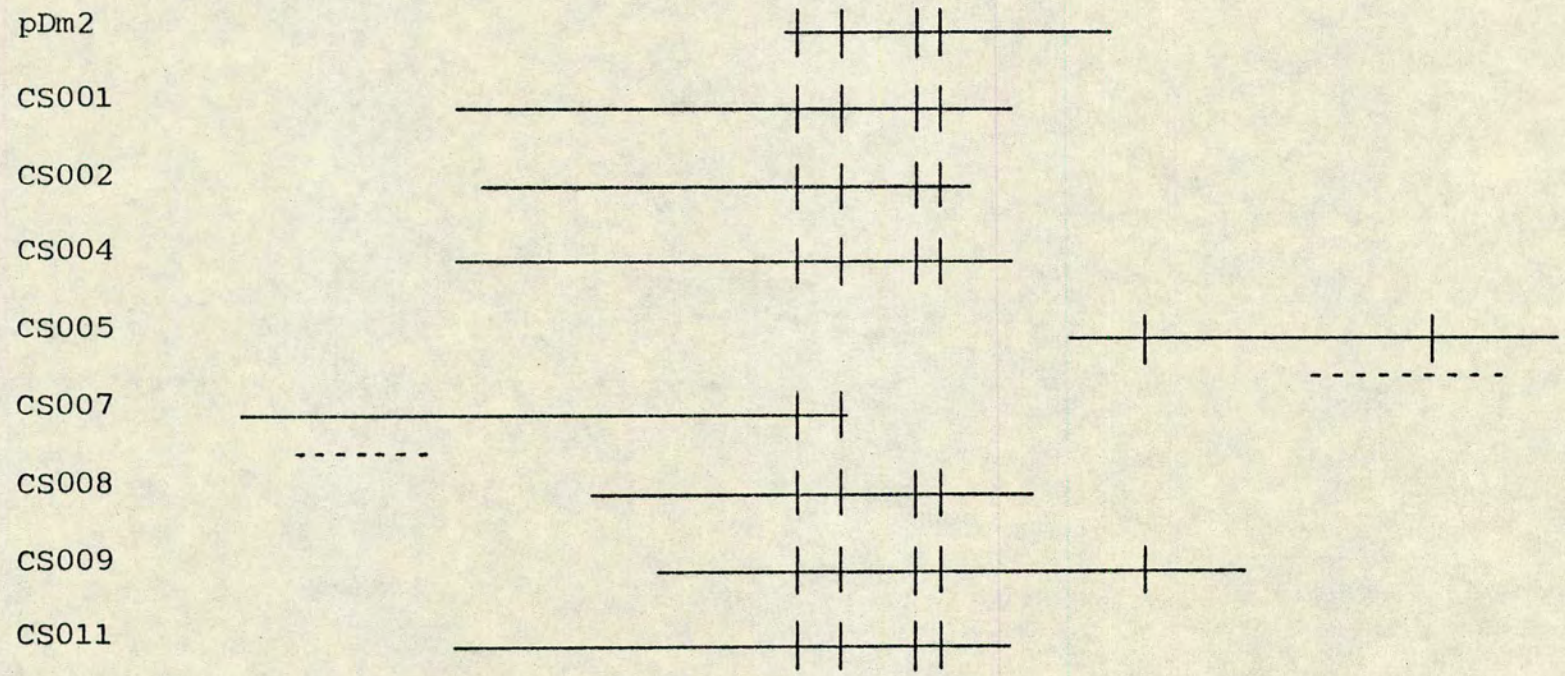
CS007 extends beyond the left end of pDm2, since it contains the 1.15 but not the 2.05kb fragments of pDm2. A small part of the 2.05kb fragment must be present however; this may be too small to appear on the gel photograph or the autoradiograph.

Finally, CS009 carries the 2.05, 1.15, and 0.55kb fragments of pDm2, a strongly hybridising 5.4kb band, a weakly hybridising 3.6kb band, and a non-hybridising band of 2.6kb. Applying the same reasoning as for CS005, CS009 must lie across pDm2, with the 3.6kb fragment extending beyond the left end, and the 5.4 and 2.6kb fragments (in that order) beyond the right end. This fixes an Eco R1 site 0.9kb to the right of the end of pDm2. This Eco R1 site must be that which gives rise to the 3kb fragment of CS005. Thus the exact start point of CS005 in pDm2 is determined. See figure 4.3.

Figure 4.3

Diagram showing regions of overlap between pDm2, CS001, CS002, CS004, CS005, CS007, CS008, CS009, and CS011. The Eco R1 sites are marked. The dashed lines indicate the restriction fragments CS005f and CS007f used as probes in the second and third screening experiments. Scale 1cm = 2kb.

Figure 4.3

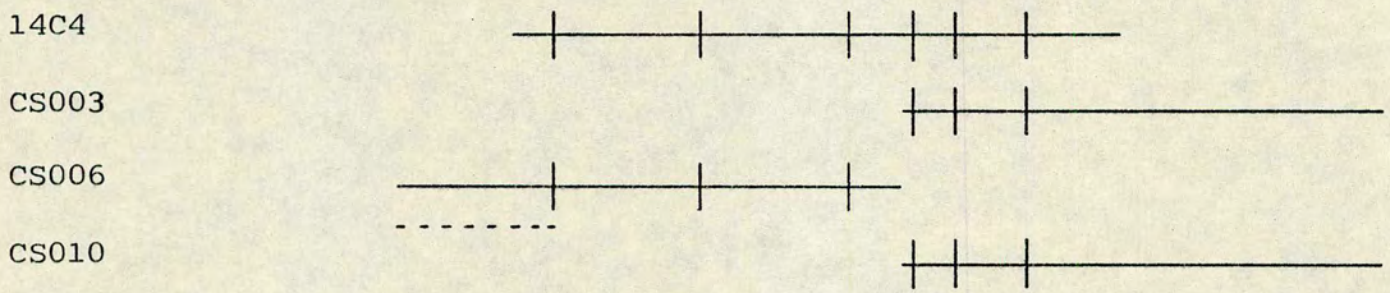


The situation with the 14C4 homologous recombinants is simpler. CS003 and CS010 are identical. They carry two fragments which hybridise with the 14C4 probe and which co-migrate with the 1.9 and 1.05kb fragments of 14C4, plus an additional hybridising band of about 9kb. The 1.7kb fragment is not present, although a short piece is presumably there which is too small to appear on the gel photograph or the autoradiograph. Hence CS003 and CS010 must extend beyond the right end of 14C4.

CS006 contains the 3.9 and 3.8kb fragments of 14C4, plus 4.2 and 1.4kb fragments which show homology with 14C4 DNA. The 1.7kb fragment of 14C4 is absent; CS006 must extend to the left of 14C4. From considerations of size, the 4.2kb fragment must extend beyond the left end, and the 1.4kb fragment be derived from the 1.7kb fragment of 14C4. See figure 4.4.

Figure 4.4
Diagram showing regions of overlap between 14C4, CS003, CS006, and CS010. The Eco R1 sites are marked. The dashed line indicates the restriction fragment CS006f used as probe in the second and third screening experiments. Scale 1cm = 2kb.

Figure 4.4



Second screening experiment.

For the second screening experiment fragments of CS005, CS006, and CS007 which had been shown not to carry repeated sequences (see section 4.5) were purified from agarose gels for use as probes. The fragments were the 5.0kb Sma 1 fragment of CS005 (CS005f, see figure 4.3), the 4.2kb Eco R1 fragment of CS006 (CS006f, see figure 4.4), and the 3.35kb Hind 111 fragment of CS007 (CS007f, see figure 4.3). No useful non-repetitive fragment could be prepared from CS003, and so 14C4 was used as probe for further recombinants in that direction. The probes prepared from the isolated fragments were contaminated to a greater or lesser extent with vector DNA, CS005f and CS007f giving significant levels of hybridisation to Charon 4 DNA, and CS006f only very faint hybridisation. In order to minimise the background hybridisation caused by the impure probes, which could mask weak positives, while maximising the number of probes with which each plaque was screened, 28 000 plaques of the CS library were screened with a mixed probe of nick translated CS005f, CS006f, and 14C4 DNAs, and 26 000 plaques, also of the CS library, with a mixed probe of nick translated CS006f, CS007f, and 14C4 DNAs. To further improve the background, denatured Charon 4 DNA was added to the hybridisation buffer to 0.6ug per ml. This approach gave 25 positively hybridising plaques of which 16 were recovered after two rounds of plaque purification (see table 4.2). This recombinants were designated CS012 to CS027. CS016, CS020, and CS021 gave very weak hybridisation.

As a first step towards characterising these recombinants, plaques of each were screened with single probes of nick translated pDm2, 14C4, CS005f, CS006f or CS007f DNA. None of the recombinants showed hybridisation above background with the pDm2, CS005f or CS007f probes. CS016 and CS020 only hybridised with 14C4, and CS019 only with CS006f. CS021 did not give detectable hybridisation with any of the probes used. The remaining recombinants

Table 4.2
Details of the second screening experiment.

Library	CS			
Probes	14C4	CS006f	CS005f	CS007f
Number of plaques screened	54 000	54 000	28 000	26 000
Probability of positive(s)[a]	100%	100%	97%	95%
Total positives expected[b]	10	7	3	3
Different positives expected[c]	10	6	3	3
Initial number of positives	(	25		)
Final number of positives	(16)		0	0
Number of different positives	(15)		0	0

Notes

(a) Calculated using equation 1.9 with the following values:-

a = 16 (Maniatis et al, 1978)
G = 165 000 (Rudkin, 1972)
N = 600 000 (Maniatis et al, 1978)
n = 54 000, 28 000, or 26 000 as appropriate.
p = 1
b = 15.85 for 14C4, 4.2 for CS006f, 5.0 for CS005f, and 3.35 for CS007f.
c = 0.2

(b) Calculated using equation 1.10 with the same values as in (a) above.

(c) Calculated using equation 1.11 with the same values as in (a) above.

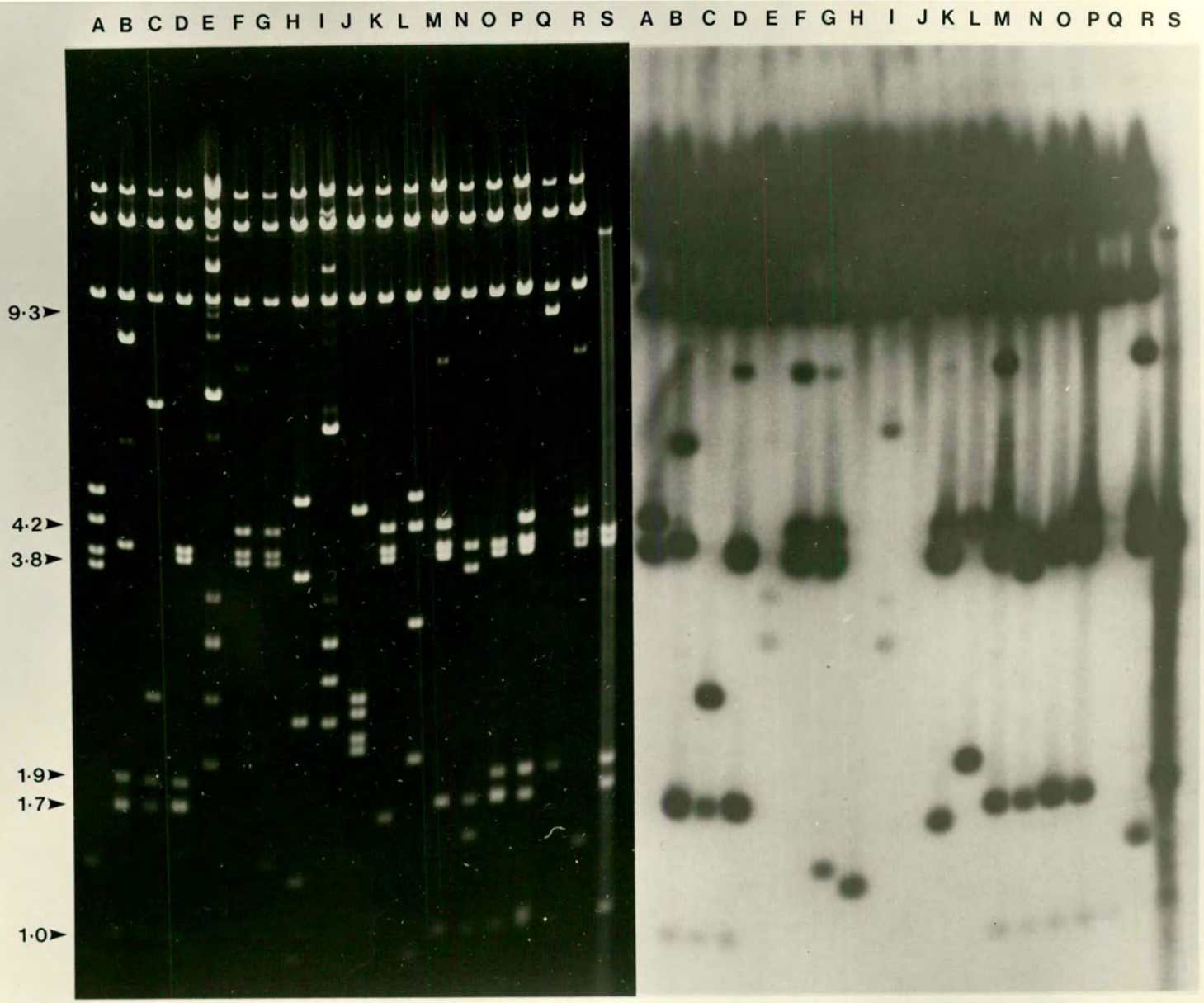
hybridised with both 14C4 and CS006f DNA.

DNA was prepared from CS012 to CS027 and digested with Eco R1, electrophoresed through 0.7% agarose gels, and the fragments transferred to nitrocellulose and probed with nick translated CS003 or CS006 DNA (see figures 4.5 and 4.6). The regions of overlap between the recombinants (see figure 4.7) were deduced using the same type of reasoning as set out for the first screening experiment. Three points should be noted. Firstly CS016, CS020, and CS021 cannot be aligned with the other recombinants. These will be discussed further in sections 4.3 and 4.6. Secondly, from the positions of overlap deduced for CS013, CS014, and CS025, they would not have been expected to have hybridised with the CS006f probe in the plaque hybridisation experiment described above. The most likely explanation for this discrepancy is that the CS006f probe was contaminated with sequences lying to the right of the 4.2kb fragment in CS006. Since Eco R1 digestion of CS006 also generates 3.9 and 3.8kb fragments this would not be unreasonable. Finally, the 1.05kb fragment of 14C4 (and several other recombinants) hybridises with the CS006 probe. This would not have been expected from the regions of overlap shown in figure 4.4, and will be considered in sections 4.3 and 4.6.

Figure 4.5

DNA from recombinants 14C4, CS003, CS006, and CS012 to CS027 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS003 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of 14C4, CS003, and CS006, giving their sizes in kb.

Track	Recombinant
A	CS012
B	CS013
C	CS014
D	CS015
E	CS016
F	CS017
G	CS018
H	CS019
I	CS020
J	CS021
K	CS022
L	CS023
M	CS024
N	CS025
O	CS026
P	CS027
Q	CS003
R	CS006
S	14C4

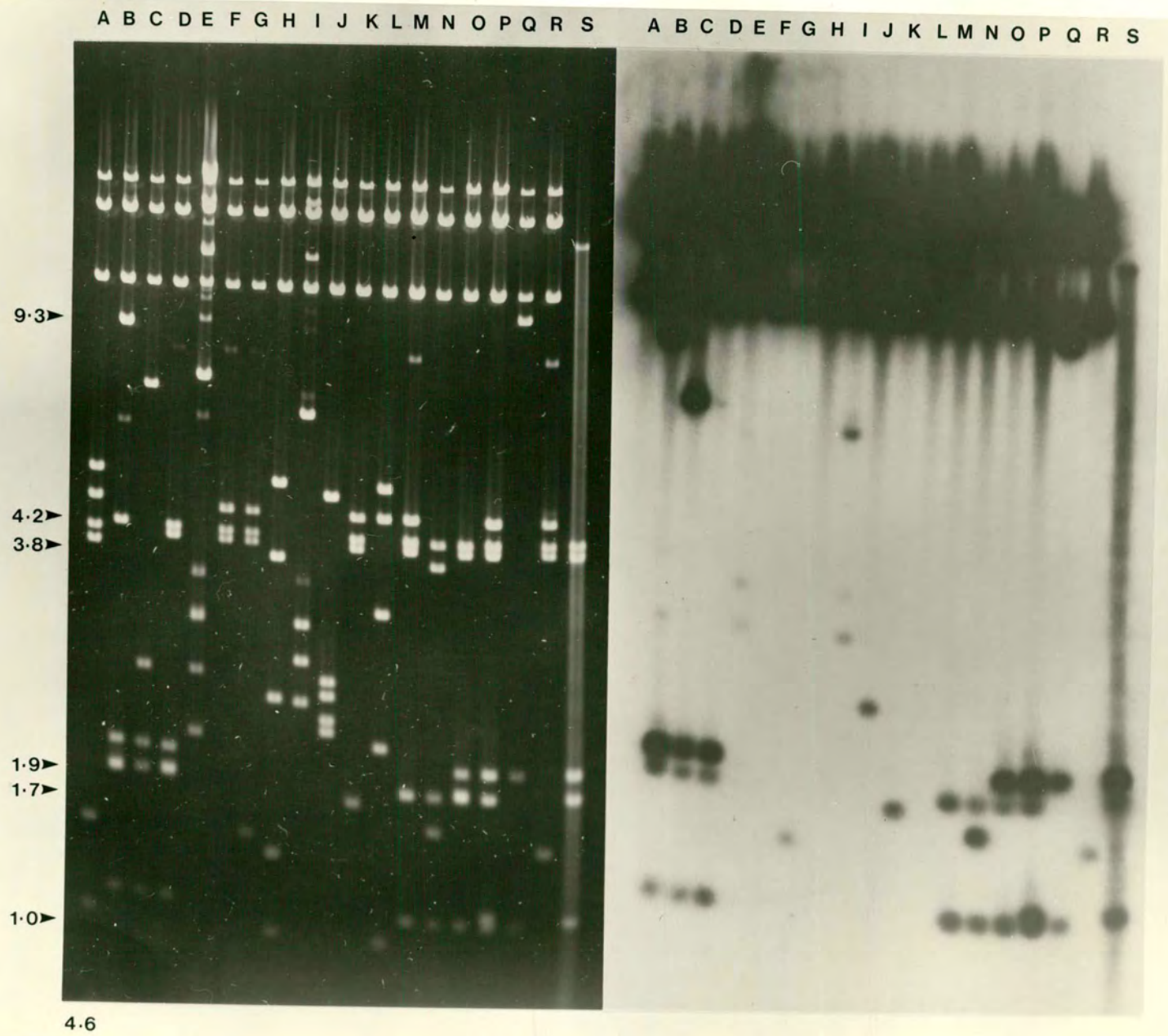

A B C D E F G H I J K L M N O P Q R S
A B C D E F G H I J K L M N O P Q R S
9·3
4·2
3·8
1·9
1·7
1·0

4·5

Figure 4.6

DNA from recombinants 14C4, CS003, CS006, and CS012 to CS027 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS006 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of 14C4, CS003, and CS006, giving their sizes in kb.

Track	Recombinant
A	CS012
B	CS013
C	CS014
D	CS015
E	CS016
F	CS017
G	CS018
H	CS019
I	CS020
J	CS021
K	CS022
L	CS023
M	CS024
N	CS025
O	CS026
P	CS027
Q	CS003
R	CS006
S	14C4

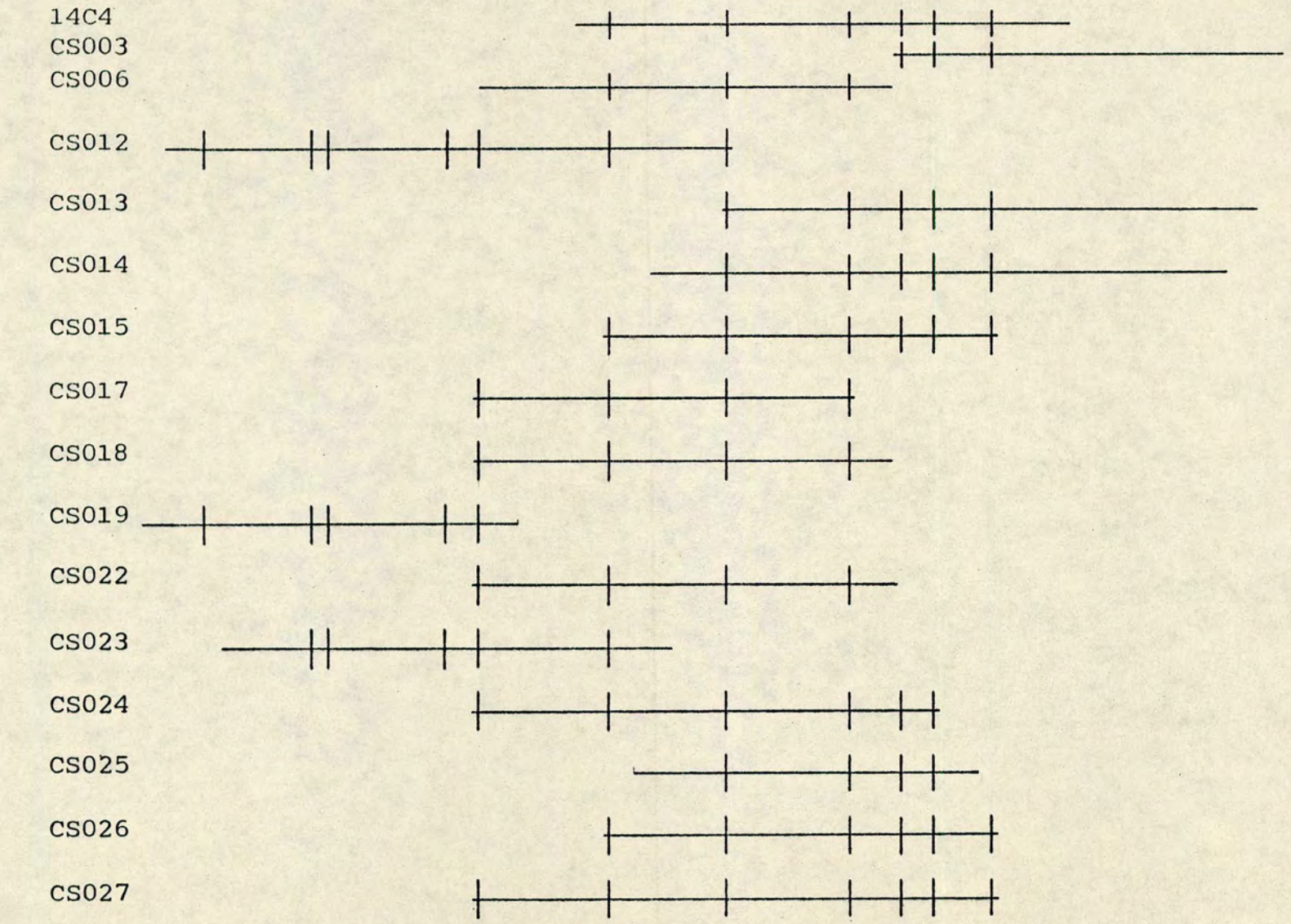

A B C D E F G H I J K L M N O P Q R S
A B C D E F G H I J K L M N O P Q R S
9·3
4·2
3·8
1·9
1·7
1·0

4.6

Figure 4.7

Diagram showing regions of overlap between 14C4, CS003, CS006, CS012, CS013, CS014, CS015, CS017, CS018, CS019, CS022, CS023, CS024, CS025, CS026 and CS027. The Eco R1 sites are marked. Recombinants CS016, CS020 and CS021 could not be aligned with the others. Scale 1cm = 2kb.

Figure 4.7


14C4
CS003
CS006
CS012
CS013
CS014
CS015
CS017
CS018
CS019
CS022
CS023
CS024
CS025
CS026
CS027

Third screening experiment.

In the third screening experiment, which was of the OR library, 24 000 plaques were probed with a mixture of nick translated pDm2, 14C4, and CS006f DNAs, 24 000 plaques with nick translated CS005f DNA, and 24 000 plaques with nick translated CS007f DNA. Initially 51 positives were identified with the mixed probe, none with the CS005f probe, and one with the CS007f probe. After two rounds of plaque purification, 29 of the positives obtained with the mixed probe still showed hybridisation, as did the positive obtained with the CS007f probe (see table 4.3). These recombinants were designated OR001 to OR029, and OR030 respectively. Plaques of OR030 did not give positive hybridisation with the mixed probe.

DNA prepared from OR001 to OR030 was digested with Eco R1, electrophoresed through 0.7% agarose gels and the fragments transferred to nitrocellulose and probed with nick translated pDm2 or 14C4 DNA (see figures 4.8 and 4.9). Selected recombinants were treated as above but either re-probed with labelled 14C4 DNA (figure 4.10) or probed with labelled CS003 (figure 4.11), CS006 (figure 4.12), CS007 (figure 4.13), or CS019 DNA (figure 4.14). Finally, nick translated OR029 and OR030 DNAs were used as probes to Eco R1 digested 14C4, CS006, and CS019 DNA, and Eco R1-Hind 111 digested CS007 DNA respectively (see figures 4.15 and 4.16). From these results the regions of overlap between the recombinants shown in figures 4.17 and 4.18 were deduced.

The recombinants OR001, OR003, OR004, OR007, and OR009 to OR027 carry the same inserts as far as can be determined. Small variations at the ends of the inserts cannot be ruled out however as the inserts cannot be cleaved from the vector DNA in their entirety, and small differences in the sizes of the large vector containing fragments would not be detected. Neither these recombinants nor OR028 can be aligned with the other

Table 4.3
Details of the third screening experiment.

Library	OR				
Probes	pDm2	14C4	CS006f	CS005f	CS007f
Number of plaques screened	(24 000)			24 000	24 000
Probability of positive(s)[a]	85%	92%	78%	79%	76%
Total positives expected[b]	3	4	3	3	2
Different positives expected[c]	2	3	2	2	1
Initial number of positives	(51)			0	1
Final number of positives	2	(26)		0	0
Number of different positives	1	(3)		0	0

Notes

(a) Calculated using equation 1.9 with the following values:-

a = 14 (average for the four recombinants isolated in this experiment
G = 165 000 (Rudkin, 1972)
N = 20 000 (M. Wolfner, personal communication)
n = 24 000
p = 1
b = 8.6 for pDm2, 15.85 for 14C4, 4.2 for CS006f, 5.0 for CS005f, and 3.35 for CS007f
c = 0.2

(b) Calculated using equation 1.10 with the same values as in (a) above.

(c) Calculated using equation 1.11 with the same values as in (a) above.

recombinants. This will be discussed further in sections 4.3 and 4.6.

0R029 and 0R030 are false positives. Note that it was necessary to prepare probes from these recombinants before this could be proven. This illustrates another of the difficulties which arise when the method used to construct a library is such that the insert cannot be completely separated from the vector sequences.

Figure 4.8

DNA from recombinants pDm2, 14C4, CS003, CS005, CS006, CS007, CS009, CS019, and OR001 to OR030 digested with Eco R1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated pDm2 DNA: gel photographs (left) and autoradiographs (right). The arrowheads indicate the positions of the three internal Eco R1 fragments of pDm2, giving their sizes in kb.

Track	Recombinant	Track	Recombinant
A	pDm2	U	CS009
B	CS003	V	CS019
C	CS005	W	not relevant
D	CS006	X	14C4
E	OR001	Y	OR016
F	OR002	Z	OR017
G	OR003	a	OR018
H	OR004	b	OR019
I	OR005	c	OR020
J	OR006	d	OR021
K	OR007	e	OR022
L	OR008	f	OR023
M	OR009	g	OR024
N	OR010	h	OR025
O	OR011	i	OR026
P	OR012	j	OR027
Q	OR013	k	OR028
R	OR014	l	OR029
S	OR015	m	OR030
T	CS007		

Notes

(1) Only those tracks giving positive hybridisation are labelled in the autoradiographs.

(2) Although the DNA in the OR019 to OR024 tracks (b to g) is degraded it can be seen that these recombinants have no homology with pDm2 DNA.

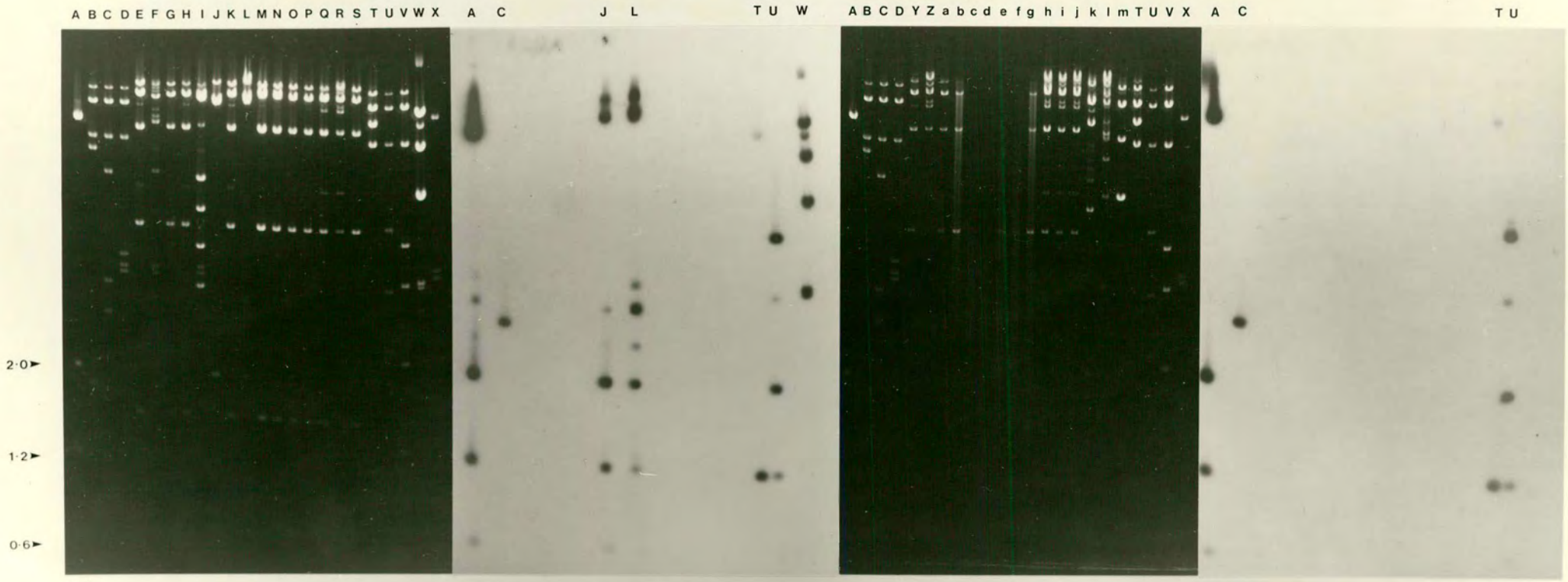

A B C D E F G H I J K L M N O P Q R S T U V W X
A C
J L
T U W
A B C D Y Z a b c d e f g h i j k l m T U V X
A C
T U
2·0
1·2
0·6
4.8

Figure 4.9

DNA from recombinants 14C4, CS003, CS005, CS006, CS007, CS009, CS019 and OR001 to OR030 digested with Eco R1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated 14C4 DNA: gel photographs (left) and autoradiographs (right). The arrowheads indicate the positions of selected fragments of 14C4, CS003, and OR001, giving their sizes in kb.

Track	Recombinant	Track	Recombinant
A	CS003	T	CS009
B	CS005	U	CS019
C	CS006	V	14C4
D	OR001	W	OR016
E	OR002	X	OR017
F	OR003	Y	OR018
G	OR004	Z	OR019
H	OR005	a	OR020
I	OR006	b	OR021
J	OR007	c	OR022
K	OR008	d	OR023
L	OR009	e	OR024
M	OR010	f	OR025
N	OR011	g	OR026
O	OR012	h	OR027
P	OR013	i	OR028
Q	OR014	j	OR029
R	OR015	k	OR030
S	CS007		

Notes

(1) The reduced level of hybridisation to the 4.9kb band of OR010, OR011, and OR012 is an artifact of this particular transfer (see figure 4.10).

(2) In the OR019 to OR024 tracks the DNA is partially degraded. However it is clear, particularly from figure 4.10, that these recombinants carry the same insert as does OR001.

(3) Digestion of OR005 DNA gives the same bands as OR002 DNA, plus several others. From the sizes and intensities of these additional bands there must be DNA from more than one phage present in the preparation of OR005 DNA. The extra bands hybridise with CS003, CS006, and CS019 DNA; this suggests that they have homology with vector rather than *Drosophila* sequences, and that therefore the contaminating phage is not from the OR library.

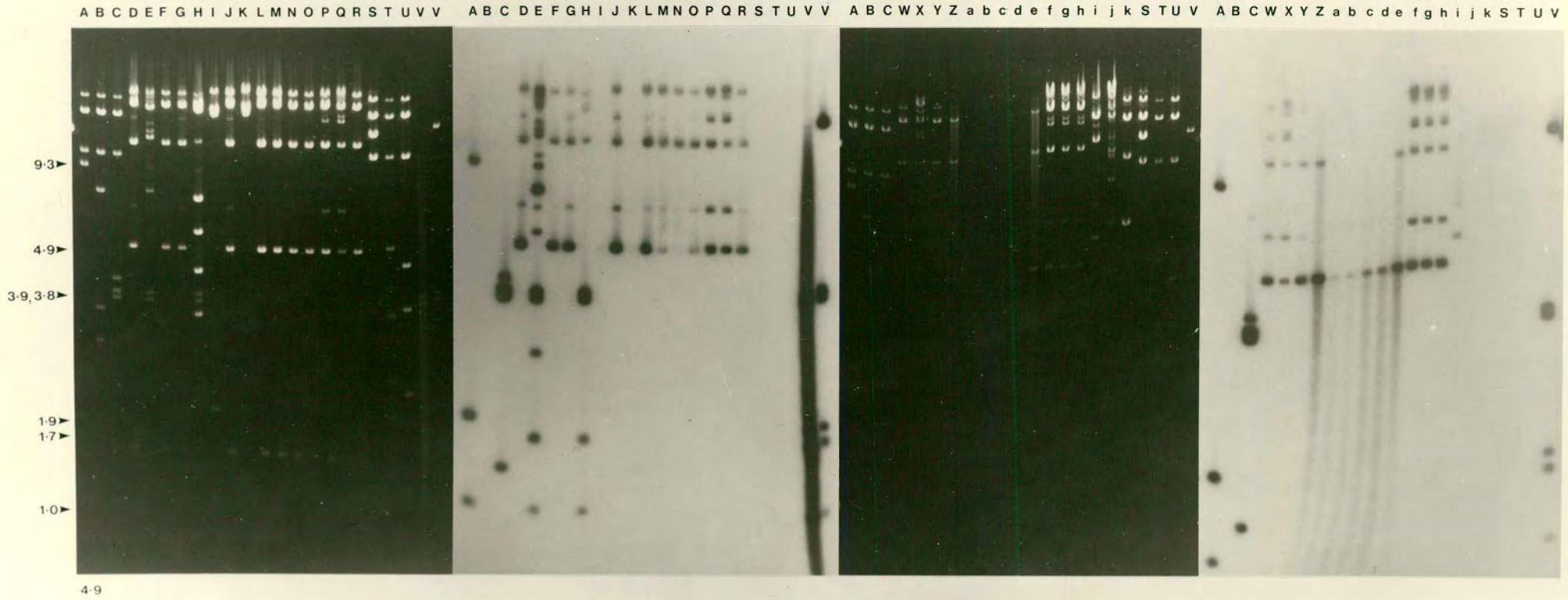
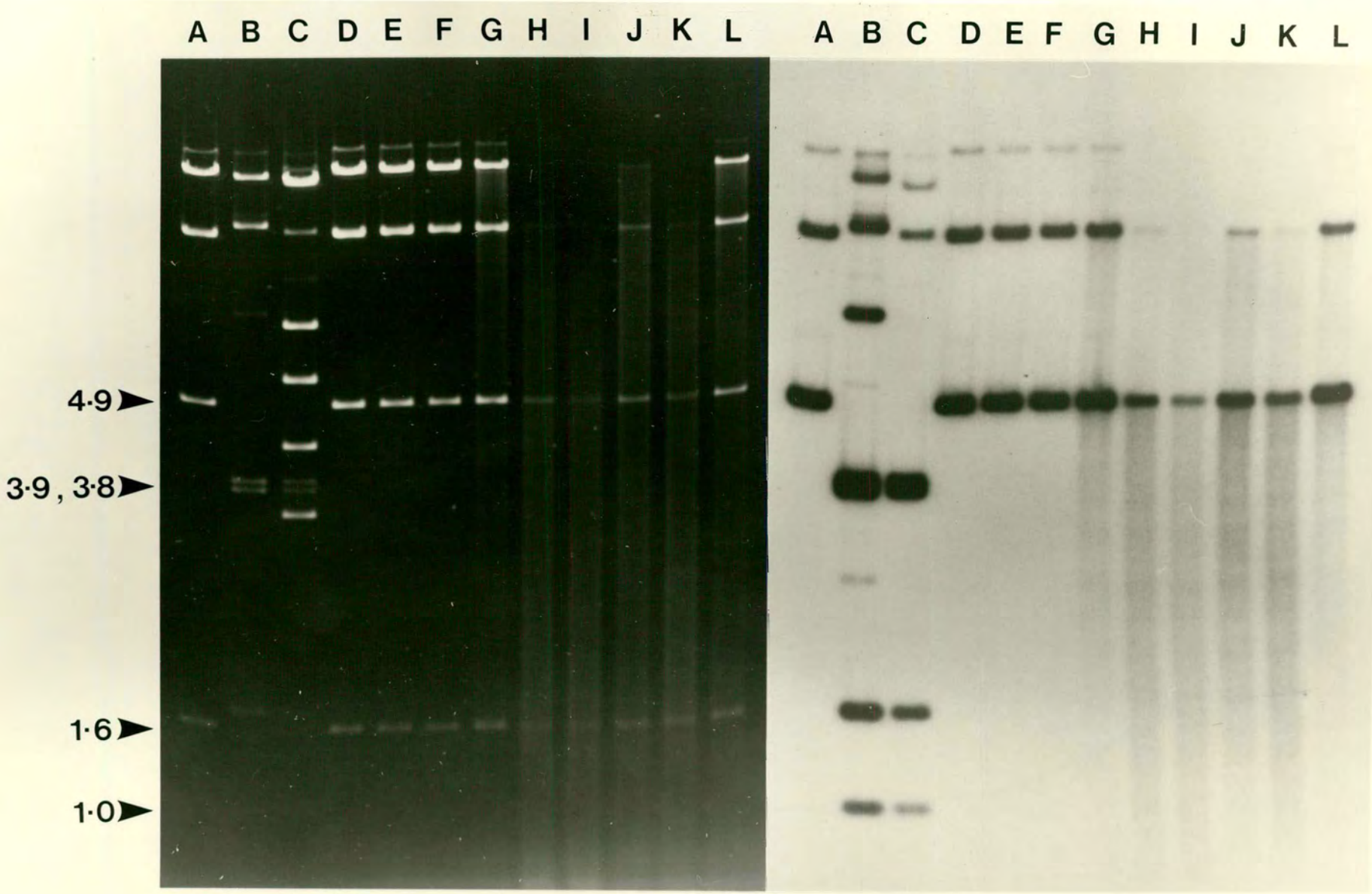

A B C D E F G H I J K L M N O P Q R S T U V V
A B C D E F G H I J K L M N O P Q R S T U V V
A B C W X Y Z a b c d e f g h i j k S T U V
A B C W X Y Z a b c d e f g h i j k S T U V
9·3
4·9
3·9, 3·8
1·9
1·7
1·0
4·9

Figure 4.10

DNA from recombinants OR001, OR002, OR005, OR010, OR011, OR012 and OR019 to OR024 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated 14C4 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of 14C4 and OR001, giving their sizes in kb.

Track	Recombinant
A	OR001
B	OR002
C	OR005
D	OR010
E	OR011
F	OR012
G	OR019
H	OR020
I	OR021
J	OR022
K	OR023
L	OR024



4·10

Figure 4.11

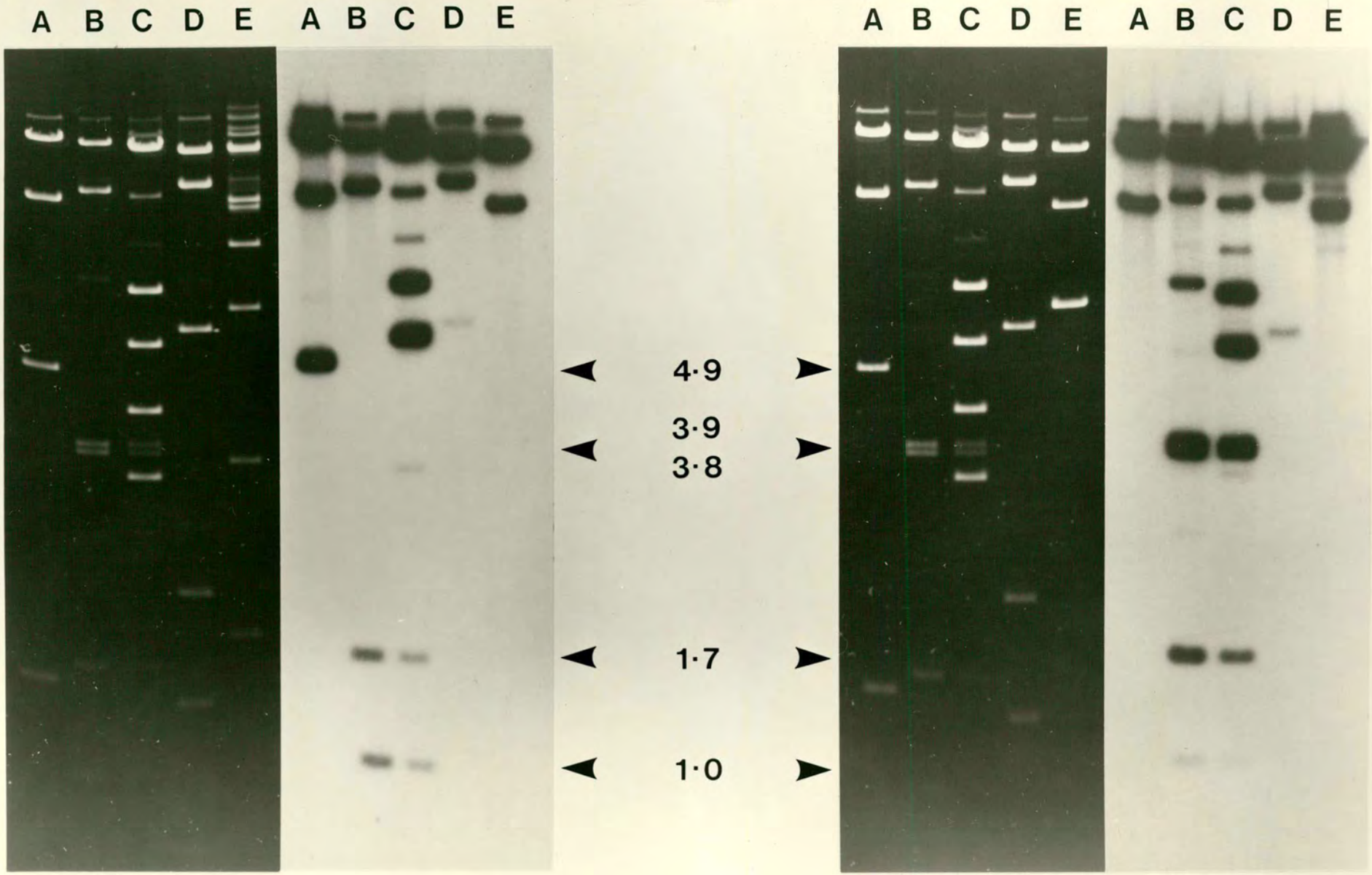
DNA from recombinants OR001, OR002, OR005, OR028 and OR030 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS003 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of 14C4, giving their sizes in kb.

Track	Recombinant
A	OR001
B	OR002
C	OR005
D	OR028
E	OR030

Figure 4.12

DNA from recombinants OR001, OR002, OR005, OR028, and OR029 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS006 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of 14C4, giving their sizes in kb.

Track	Recombinant
A	OR001
B	OR002
C	OR005
D	OR028
E	OR029



4.11

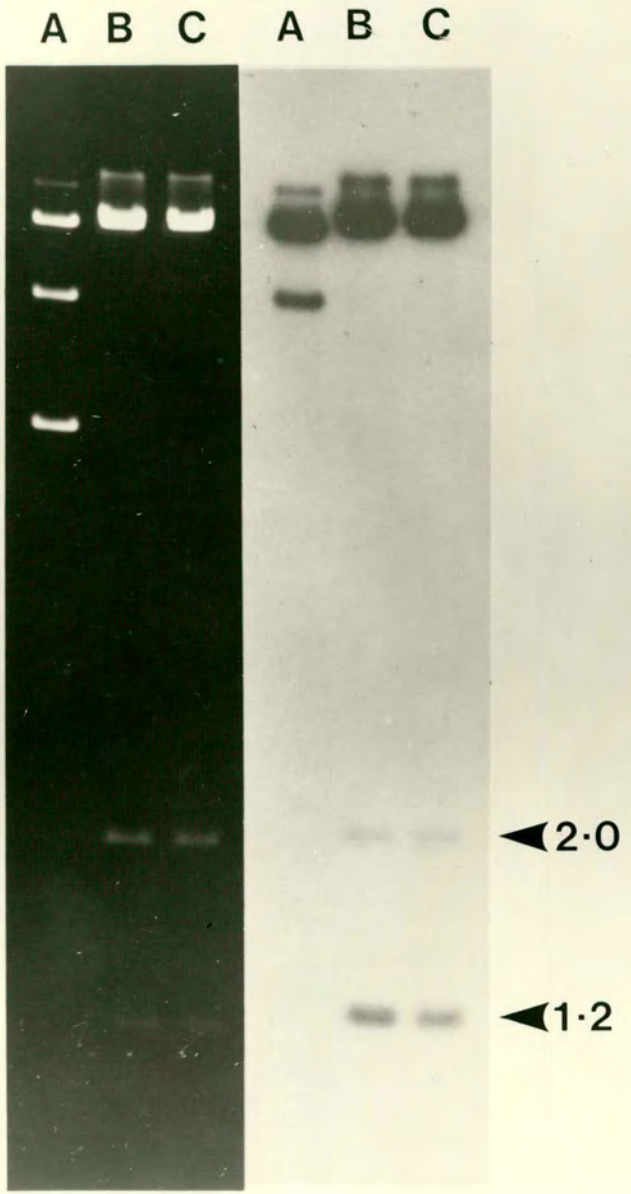
4.12

Figure 4.13
DNA from recombinants OR006, OR008, and OR030 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS007 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of pDm2, giving their sizes in kb.

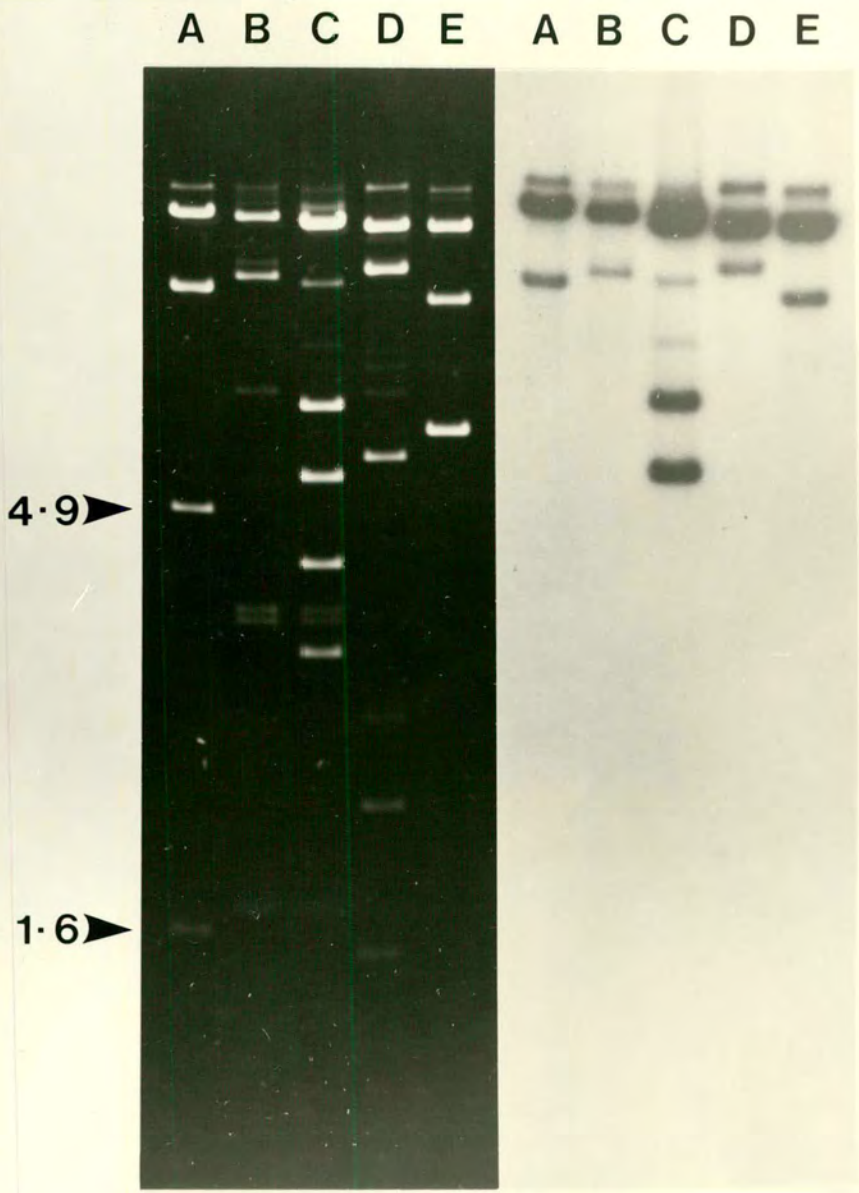
Track	Recombinant
A	OR030
B	OR006
C	OR008

Figure 4.14
DNA from recombinants OR001, OR002, OR005, OR028, and OR030 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS019 DNA: gel photograph (left) and autoradiograph (right). The arrowheads indicate the positions of selected fragments of OR001, giving their sizes in kb.

Track	Recombinant
A	OR001
B	OR002
C	OR005
D	OR028
E	OR030



4·13



4·14

Figure 4.15

DNA from recombinants 14C4, CS006, CS019, and OR029 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated OR029 DNA: gel photograph (left) and autoradiograph (right).

Track	Recombinant
A	14C4
B	CS006
C	CS019
D	OR029

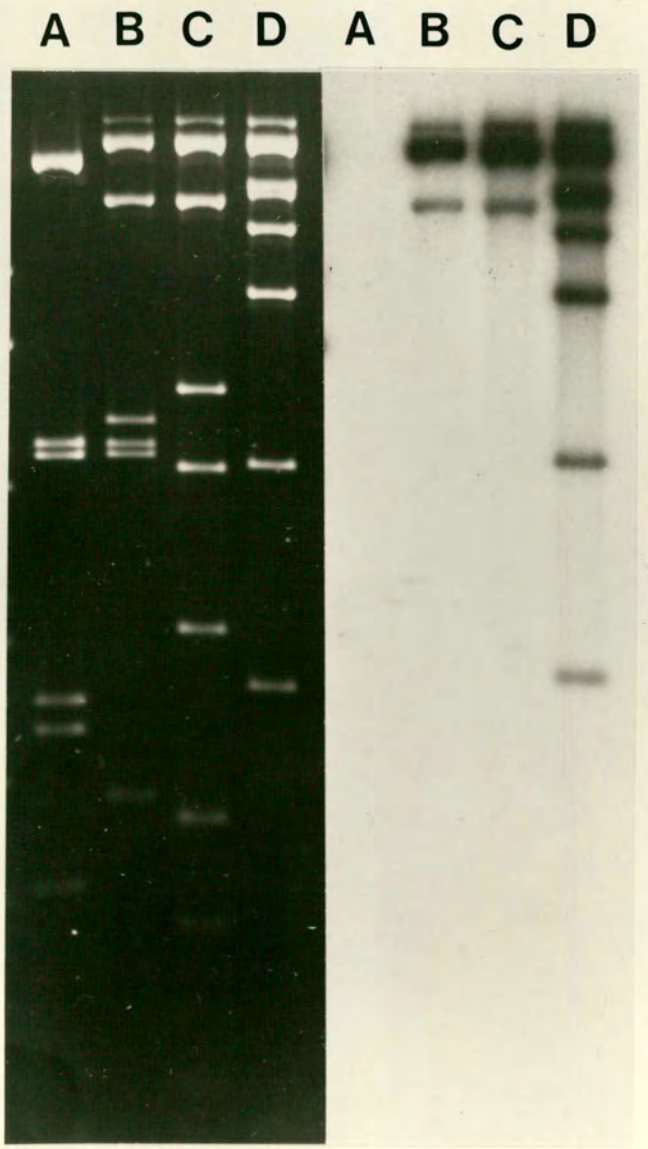
Figure 4.16

DNA from recombinants CS007 and OR030 digested with Eco R1 and Hind 111, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated OR030 DNA: gel photograph (left) and autoradiograph (right).

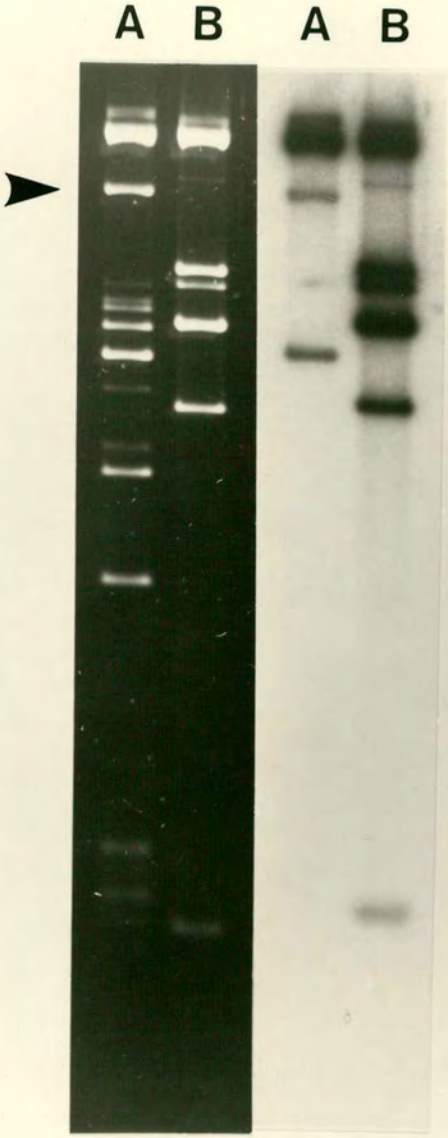
Track	Recombinant
A	CS007
B	OR030

Notes

(1) The band marked with an arrowhead is a product of partial digestion and hybridises by virtue of vector homology.



4·15



4·16

Figure 4.17

Diagram showing regions of overlap between pDm2, CS005, CS007, CS009, OR006 and OR008. The Eco R1 sites are marked. Note that the exact end points of the OR recombinants cannot be determined without detailed restriction mapping (see section 4.2). Scale 1cm = 2kb.

Figure 4.17

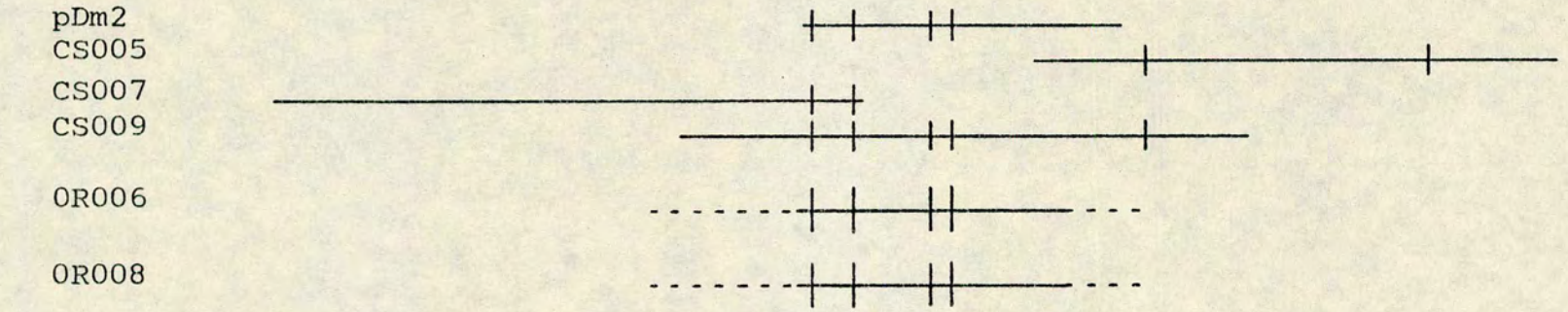
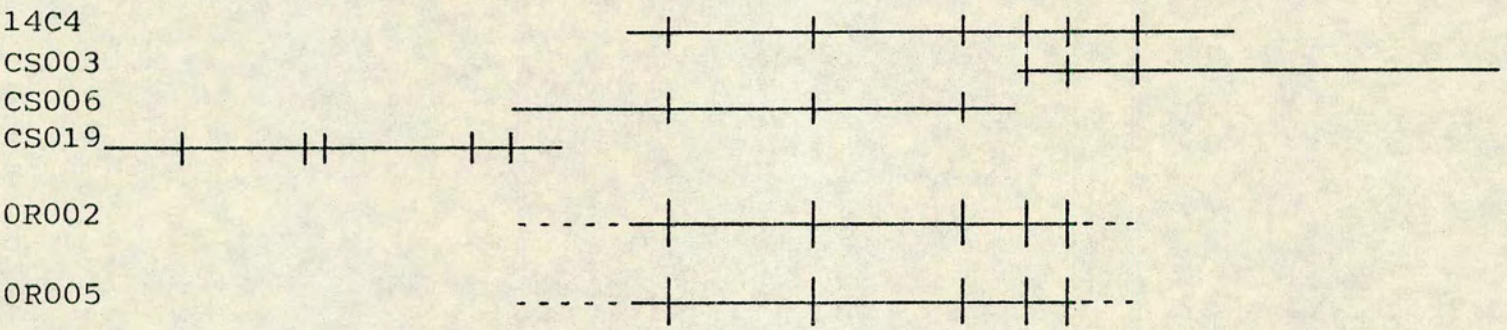


Figure 4.18

Diagram showing regions of overlap between 14C4, CS003, CS006, CS019, OR002 and OR005. The Eco R1 sites are marked. Note that the exact end points of the OR recombinants cannot be determined without detailed restriction mapping (see section 4.2). Recombinants OR001, OR003, OR004, OR007, OR009 to OR027 (all of which carry identical inserts (see text)) and OR028 could not be aligned with the others. Scale 1cm = 2kb.

Figure 4.18



Discussion.

The deviations between the predicted number of positives and the actual numbers obtained in the screening experiments may be due to effects of amplification and differential viability (as discussed in section 1) as well as simple statistical fluctuations. The exact histories of the CS and OR libraries are not known, but the CS library has been amplified at least once, and the OR library at least twice. For simplicity it has been assumed in tables 4.1, 4.2, and 4.3 that the libraries have only been amplified once, and that the entire library was amplified in each case. In view of this the results obtained seem reasonable, with the possible exceptions of the 23 OR001-like recombinants recovered in the third screening, and the failure to pick up any positives with the CS005f and CS007f probes.

The OR001-like phage did not show any obvious difference in viability from the other OR recombinants recovered; hence increased viability is unlikely to be the sole explanation, although it could be a contributing factor. It must therefore be concluded that either the effects of amplification are much more severe than anticipated, or that this is a rare chance occurrence.

The most likely explanation for the failure to identify recombinants having homology with the CS005f and CS007f probes is their low specific activity. Fragments isolated from gels were poor substrates for nick translation. As the exact concentrations of the DNAs were not determined, the specific activities cannot be calculated accurately: however they are estimated to vary between one-half and one-twentieth of the specific activities attained with the plasmid probes (pDm2 and 14C4). Thus the minimum detectable homology will be greater and the probability of isolating recombinants reduced. The recombinants CS016 and CS020, which are believed to share about 0.2kb of sequence with 14C4 (see sections 4.3 and 4.6), were identified using a mixed probe containing 14C4, CS005f, and CS006f DNAs.

The raised background due to contamination of CS005f with vector DNA is therefore not important. The background given with the CS007f probe was slightly higher and may have been a minor contributing factor in this case.

The loss of positives during plaque purification is mainly due to artifacts of the plaque hybridisation producing false positives. Maniatis et al (1978) found that 88% of the positives detected initially still gave hybridisation on re-screening. The corresponding figures for the first, second, and third screening experiments are 74%, 64%, and 54% respectively. This presumably reflects the increasing care taken in picking up very weak positives. Note however that even plaques which initially gave quite strong hybridisation may be negative on re-screening.

On the basis of the results shown in figures 4.3, 4.4, 4.7, 4.17, and 4.18, those recombinants which extend the "walk" furthest in both strains were selected for more detailed examination. These recombinants were pDm2, CS005, CS007, CS009, and OR006 (pDm 2 related group), 14C4, CS003, CS006, CS019, and OR002 (14C4 related group), and CS016, CS020, CS021, OR001, and OR028 (having homology with members of the 14C4 related group, but being incapable of alignment with the members of that group). In determining the regions of overlap it has been temporarily assumed that there were no low frequency repeated sequences on any of the probes, and that none of the recombinants carry double inserts or alternative arrangements of sequence (see section 1). These possibilities will be considered fully in section 4.7.

4.2 Restriction Mapping.

The restriction maps presented in this section were deduced from the sizes of the fragments produced by single, double, and, occasionally, triple digestion of purified DNA with selected restriction enzymes. The sizes of the fragments were determined by electrophoresis through agarose and polyacrylamide gels of appropriate concentrations and comparison with fragments of known molecular weight run on the same gels. Fragment sizes determined by this method varied by up to 10% between different gels. The sizes given in the maps are the averages of the values obtained.

In the maps of the phage recombinants only the inserts are shown. Maps of the vector arms are given separately in figure 4.33 (Charon 4, vector for the CS library) and figure 4.34 (Sep 6, vector for the OR library). In all cases digestion of the recombinant DNA produced the expected vector derived fragments. The maps are drawn with the junction with the left vector arm at the left. The maps of pDm2 and 14C4 include the vectors. Here the orientations are purely arbitrary.

The data from which these maps were constructed will not be presented. However, the limit of resolution for each map (the smallest fragment which would have been detected on the gels used) is indicated in the figure legend, as is any additional information used to generate that particular map.

The Eco R1 fragments (or, in the case of CS007 and OR006, the Hind 111 fragments) are labelled alphabetically in order of size for ease of reference in later sections.

Restriction sites are abbreviated as follows:-

b - Bam H1
c - Sac 1
e - Eco R1
h - Hind 111
k - Kpn 1
l - Sal 1
p - Pst 1
s - Sma 1
x - Xho 1

All the maps in this section are drawn to a scale of 1cm = 1kb.

Figure 19.
Restriction map of pDm2.

Notes

(1) The limit of resolution of the map is 0.08kb, except for Sma 1, where it is 0.6kb.

(2) The Eco R1 fragment order was confirmed by analysis of the fragment sizes produced by partial digestion with Eco R1.

(3) The positions of the junctions between insert and vector shown opposite were determined from the distance between the Hind 111 site in pSC101 and the Eco R1 site into which the insert was cloned (Boyer et al, 1977).

Figure 19.
Restriction map of pDm2.

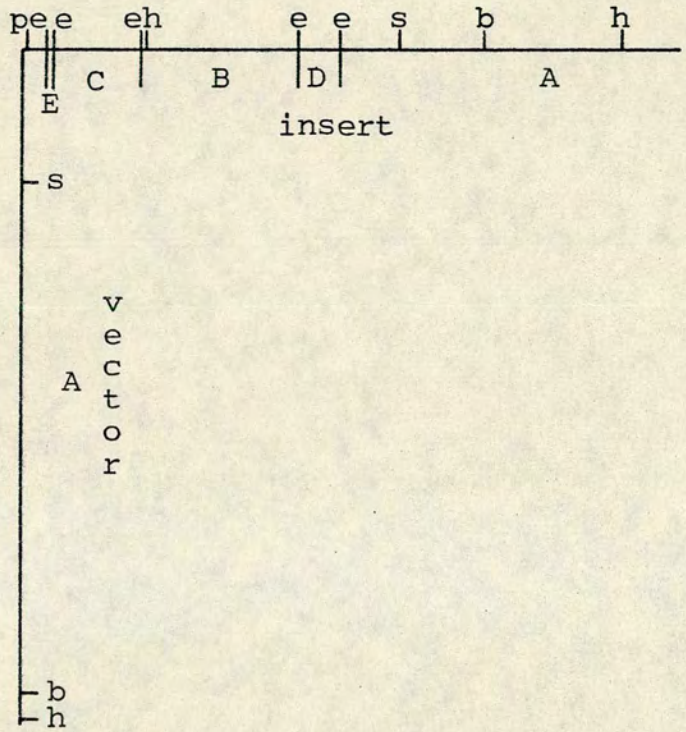


Figure 20.
Restriction map of 14C4.

Notes

(1) The limit of resolution of the map is 0.1kb.

(2) Those fragments which contain vector sequences were identified by transfer of agarose gels of restriction digests to nitrocellulose and probing with nick translated RSF2124 DNA. This facilitated construction of the map considerably.

(3) The positions of the junctions between insert and vector shown opposite were determined from the distances between the Pst 1 sites in RSF2124 and the Eco R1 site into which the insert was cloned (D. Sherrat, personal communication). The map of the vector shown here is in reasonable agreement with that obtained by D. Sherrat.

(4) There are five sites for Hpa 1 within 14C4, resolution 0.1kb.

(5) The length of 14C4 determined by electron microscopy was 27.95kb ($\pm$ 0.66). This is in good agreement with the value of 26.9kb obtained by restriction mapping.

(6) The map shown here differs considerably from that published by Dudler et al (1980). See section 5.

Figure 20.
Restriction map of 14C4.

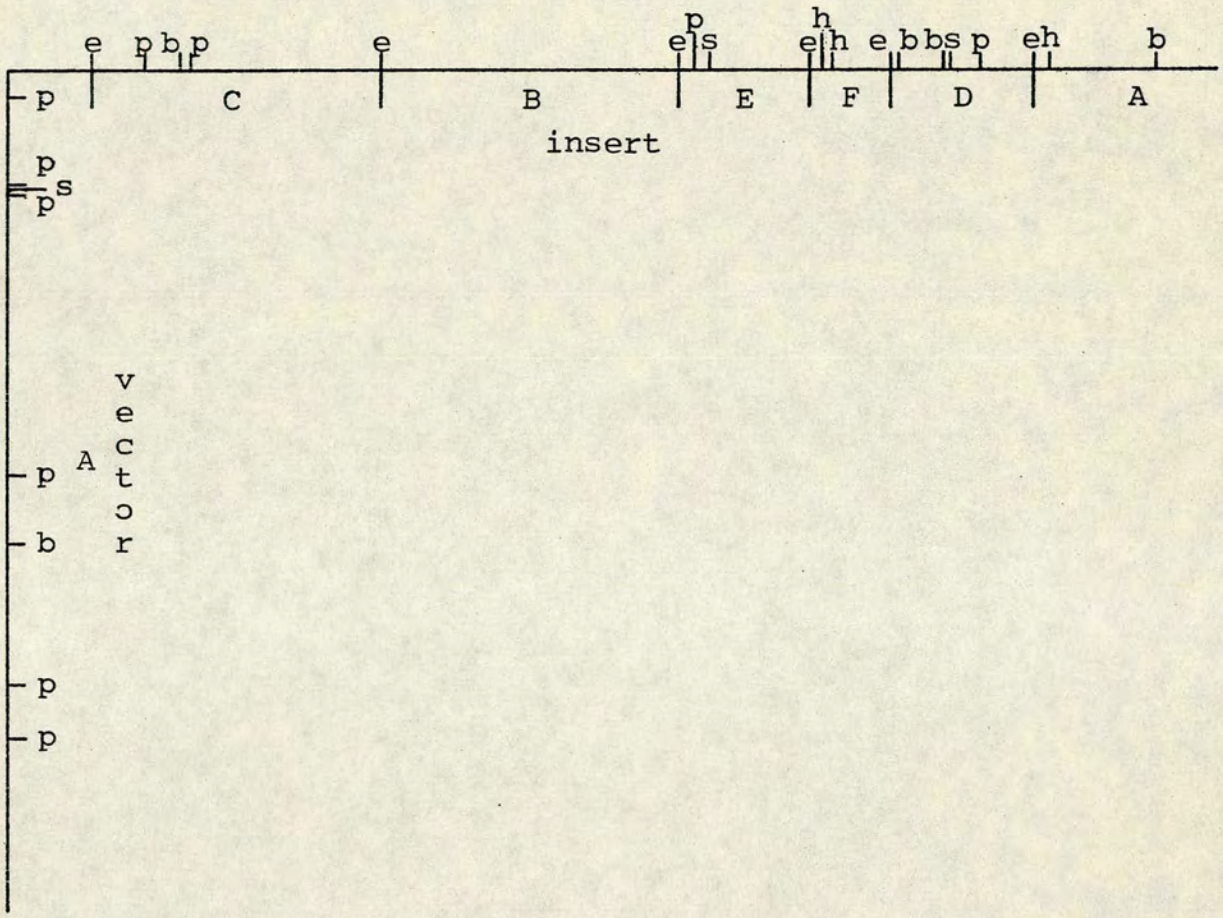
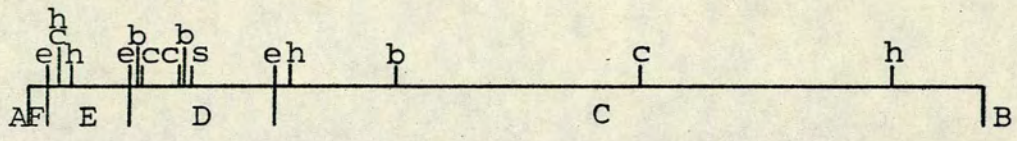


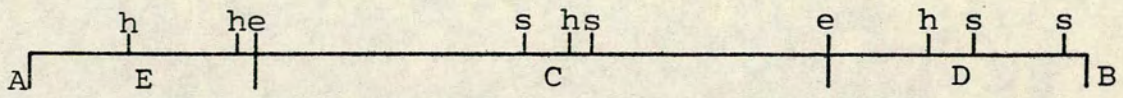
Figure 21.
Restriction map of CS003.



Notes

(1) The limit of resolution of the map is 0.1kb, except for Sma 1 where it is 0.3kb.

Figure 22.
Restriction map of CS005.

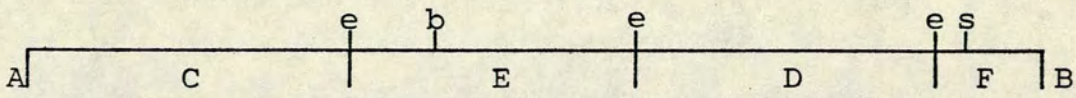


Notes

(1) The limit of resolution if the map is 0.1kb.

(2) There are no sites for Bam H1 within the insert of CS005.

Figure 23.
Restriction map of CS006.



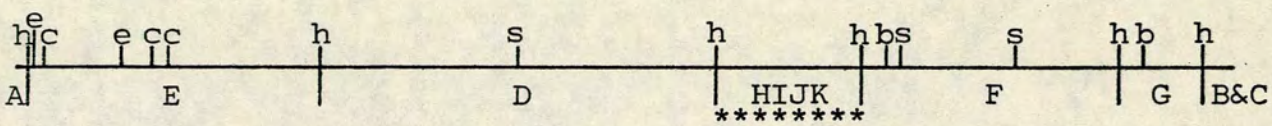
Notes

(1) The limit of resolution of the map is 0.1kb.

(2) There are no sites for Hind 111 within the insert of CS006.

(3) The 4.2 and 3.9kb Eco R1 fragments could not be placed unambiguously using only the data obtained by restriction digestion. They are assigned to the positions shown on the basis of cross-hybridisation experiments (see section 4.3) and by comparison with the maps of 14C4 (figure 20) and OR002 (figure 30).

Figure 24.
Restriction map of CS007.



Notes

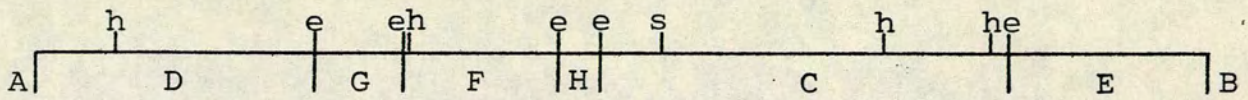
(1) The limit of resolution of the map is 0.1kb.

(2) There are three sites within the insert of CS007 which it has not been possible to map. These generate fragments of length 0.95, 0.45, 0.3, and 0.2kb. At least one of these sites, generating the 0.95 and 0.45kb fragments must lie within the region marked ***** and, on considerations of size alone, all three may lie in here.

(3) The positions of the sites for Sac 1 shown above are based solely upon one Eco R1-Sac 1 double digest and have not been confirmed with other enzymes.

(4) The length of CS007 determined by electron microscopy was 46.64kb (± 0.53). This is in good agreement with the value of 46.0kb obtained by restriction mapping.

Figure 25.
Restriction map of CS009.

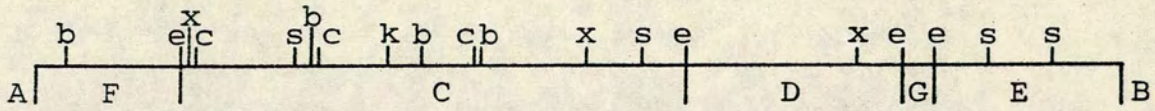


Notes

(1) The limit of resolution of the map is 0.1kb.

(2) There are no sites for Bam H1 within the insert of CS009.

Figure 26.
Restriction map of CS016.

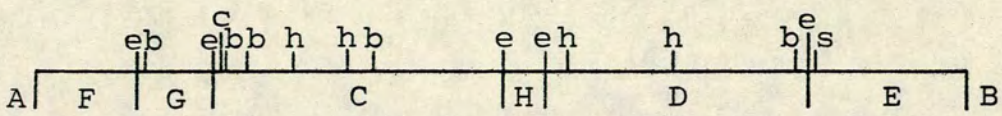


Notes

(1) The limit of resolution of the map is 0.3kb.

(2) There are eight sites for Hind 111 within the insert of CS016, resolution 0.3kb.

Figure 27.
Restriction map of CS019.



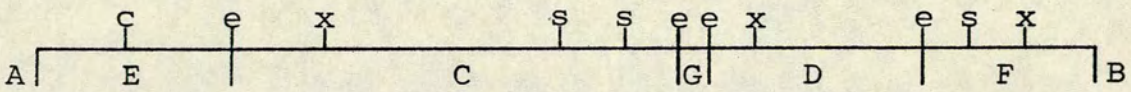
Notes

(1) The limit of resolution of the map is 0.3kb.

(2) In order to construct this map it was necessary to make use of the order of Eco R1 fragments indicated in figure 4.7. This information was derived from the results of cross-hybridisation with CS006 (see also section 4.3) and comparison of the fragments carried by CS012, CS019, and CS023.

(3) Although the map is unambiguous, the C fragment, which when measured itself has a length of about 4.8kb, appears to be composed of fragments having a total length of only 3.8kb. The reason for this large discrepancy is not known.

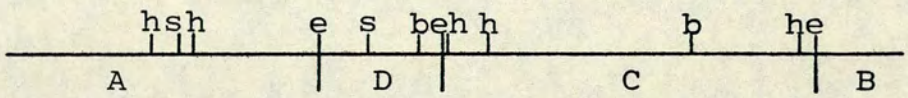
Figure 28.
Restriction map of CS020.



Notes

(1) The limit of resolution of the map is 0.3kb.
(2) There are no sites for Bam H1 within the insert of CS020.
(3) There are no sites for Kpn 1 within the insert of CS020.
(4) There are ten sites for Hind 111 within the insert of CS020, resolution 0.3kb.

Figure 29.
Restriction map of OR001.



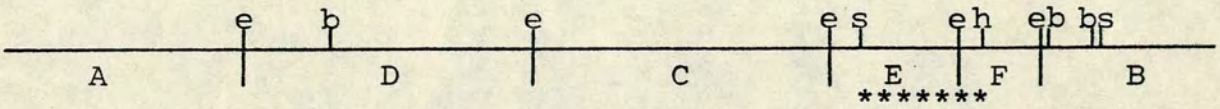
Notes

(1) The limit of resolution of the map is 0.3kb.

(2) There are no sites for Kpn 1 within the insert of OR001.

(3) The positions of the junctions between insert and vector shown above were determined from the distances between the Kpn 1 and Sma 1 sites in Sep 6 and the Eco R1 sites into which the insert was cloned (M. Wolfner, personal communication, see also figure 4.34).

Figure 30.
Restriction map of OR002.



Notes

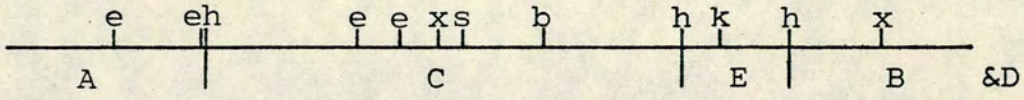
(1) The limit of resolution of the map is 0.3kb.

(2) There are no sites for Kpn 1 within the insert of OR002.

(3) The positions of the junctions between insert and vector shown above were determined from the distances between the Kpn 1 and Sma 1 sites in Sep 6 and the Eco R1 sites into which the insert was cloned (M. Wolfner, personal communication, see also figure 4.34).

(4) The fragment marked ***** is expected to be 1.6kb long but in practice is found to be only 1.4kb. The expected 0.3kb Eco R1-Hind 111 fragment, although just within the limit of resolution, is absent. In both 14C4 and CS003 there is an additional Hind 111 site 0.15kb to the left of that marked here (see figures 4.20 and 4.21). If this were present in OR002 it would account for these discrepancies.

Figure 31.
Restriction map of OR006.

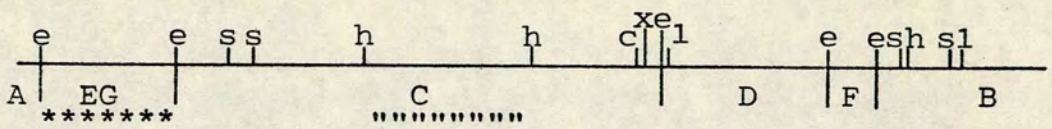


Notes

(1) The limit of resolution of the map is 0.3kb.

(2) The positions of the junctions between insert and vector shown above were determined from the distances between the Kpn 1 and Xho 1 sites in Sep 6 and the Eco R1 sites into which the insert was cloned (M. Wolfner, personal communication, see also figure 4.34).

Figure 32.
Restriction map of OR028.



Notes

(1) The limit of resolution of the map is 0.3kb.

(2) There are no sites for Bam H1 within the insert of OR028.

(3) There are no sites for Kpn 1 within the insert of OR028.

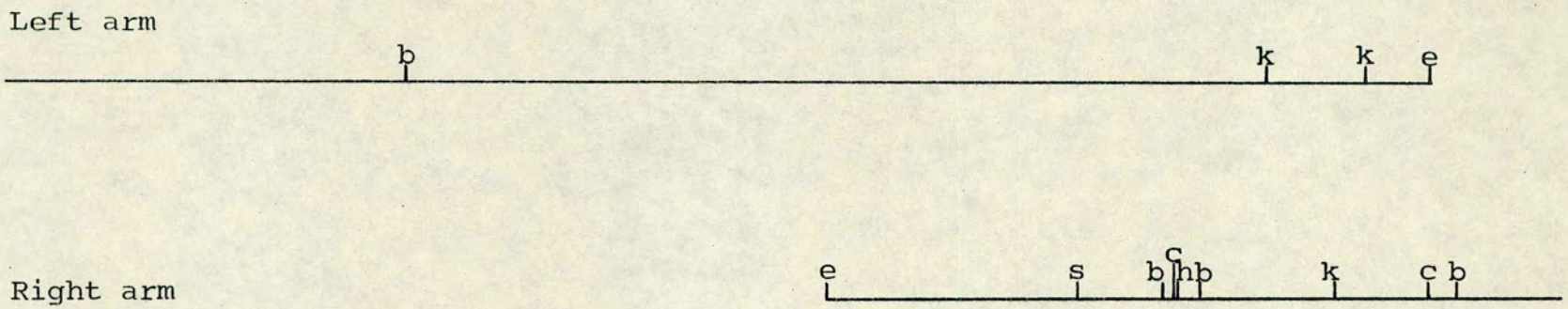
(4) The positions of the junctions between insert and vector shown above were determined from the distances between the Kpn 1 and Sal 1 sites in Sep 6 and the Eco R1 sites into which the insert was cloned (M. Wolfner, personal communication, see also figure 4.34).

(5) There is an additional Eco R1 site in the region marked ***** which cannot be placed unambiguously. This site gives rise to fragments of 1.4 and 0.4kb.

(6) There is an additional Hind 111 site in the region marked """""" which cannot be placed unambiguously. This site gives rise to fragments of 1.45 and 0.75kb.

Figure 33.

Restriction map of the arms of vector Charon 4.



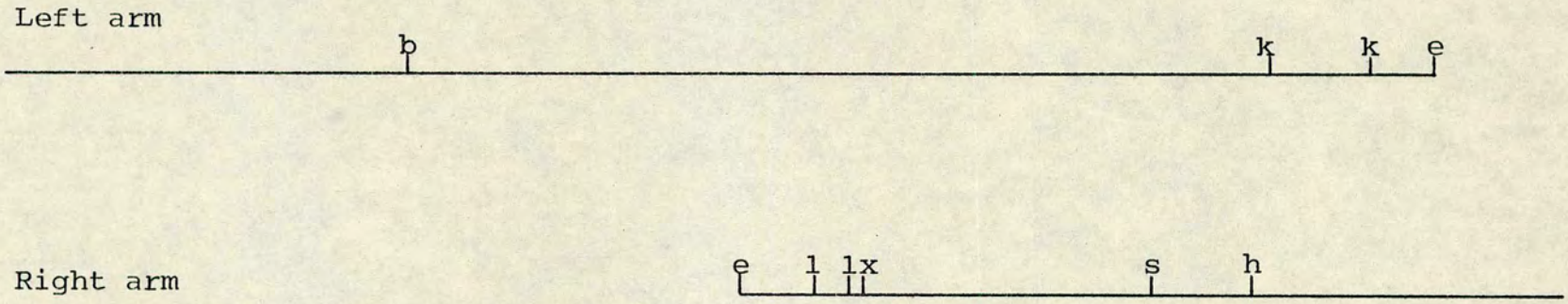
Notes

(1) The limit of resolution of the map is 0.3kb.

(2) There are no sites for Xho 1 within the vector arms.

(3) In the CS library the inserts are attached to the Eco R1 sites using Eco R1 linkers. These sites are therefore present in the CS recombinants.

Figure 34.
Restriction map of the arms of vector Sep 6.



Notes

(1) The limit of resolution of the map is 0.3kb.

(2) There are no sites for Sac 1 within the vector arms.

(3) This map is in good agreement with that of M. Wolfner (personal communication).

(4) In the OR library the inserts are tailed into the Eco R1 sites. These sites are therefore not present in the OR recombinants.

Discussion.

The restriction maps shown in figures 4.19 to 4.32 will now be used to construct maps of the regions of Drosophila melanogaster cloned. The possibility that these maps are not representative of a common arrangement of sequence in embryos of the appropriate strain will be considered in section 4.7 after all the relevant data have been presented.

The pDm2 group.

This consists of pDm2, CS005, CS007, CS009, and OR006.

Both pDm2 and OR006 contain inserts originally derived from Oregon R strain embryos. Their restriction maps overlap in the manner shown in figure 4.35. The leftmost Eco R1 site in pDm2 (which generates the 0.1kb E fragment) would not have been detected in the experiments by which the other recombinants were mapped. This site will not be considered further and is omitted from figure 4.35. Given this, the restriction maps of pDm2 and OR006 are in perfect agreement.

From the restriction maps of CS005, CS007, and CS009, the inserts of which are all from Canton S embryo DNA, these overlap as shown in figure 4.35. Again there are no discrepancies.

Now consider the two strains. The map of the region from the Oregon R strain and the map of the region from the Canton S strain show eight out of ten restriction sites in common. The two differences are a site for Bam H1 in pDm2 and OR006 which is absent from CS009, and a site for Eco R1 in CS005 and CS009 which is absent from OR006. It is not possible to say whether these represent strain specific differences or simply heterogeneity within the strains.

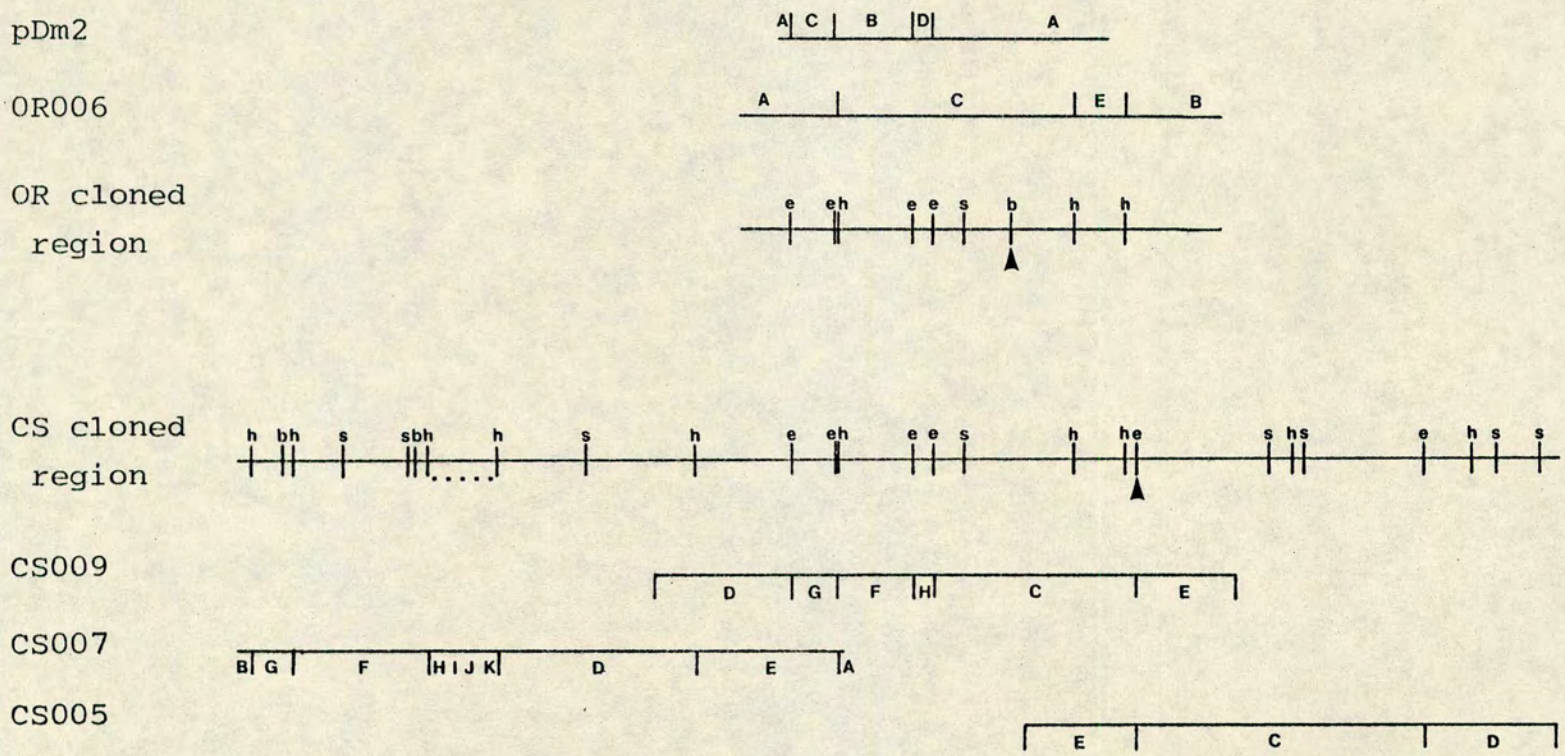
The 14C4 group.

This consists of 14C4, CS003, CS006, CS019, and OR002.

Figure 4.35

Restriction maps of the regions of the genome from Drosophila melanogaster strains Canton S (CS) and Oregon R (OR) represented by the pDm2 group of recombinants (pDm2, CS005, CS007, CS009, and OR006). Sites for the enzymes Bam H1 (b), Eco R1 (e), Hind 111 (h) and Sma 1 (s) are shown. The regions represented by the individual recombinants are indicated, and the Eco R1 fragments of each are labelled alphabetically in decreasing order of size (except in the cases of CS007 and OR006 where the Hind 111 fragments are labelled). Arrowheads mark sites present in the cloned region from one strain but not the other. There are three Hind 111 sites within the dotted region which have not been mapped (see figure 4.24). Scale 1cm = 2kb.

Figure 4.35

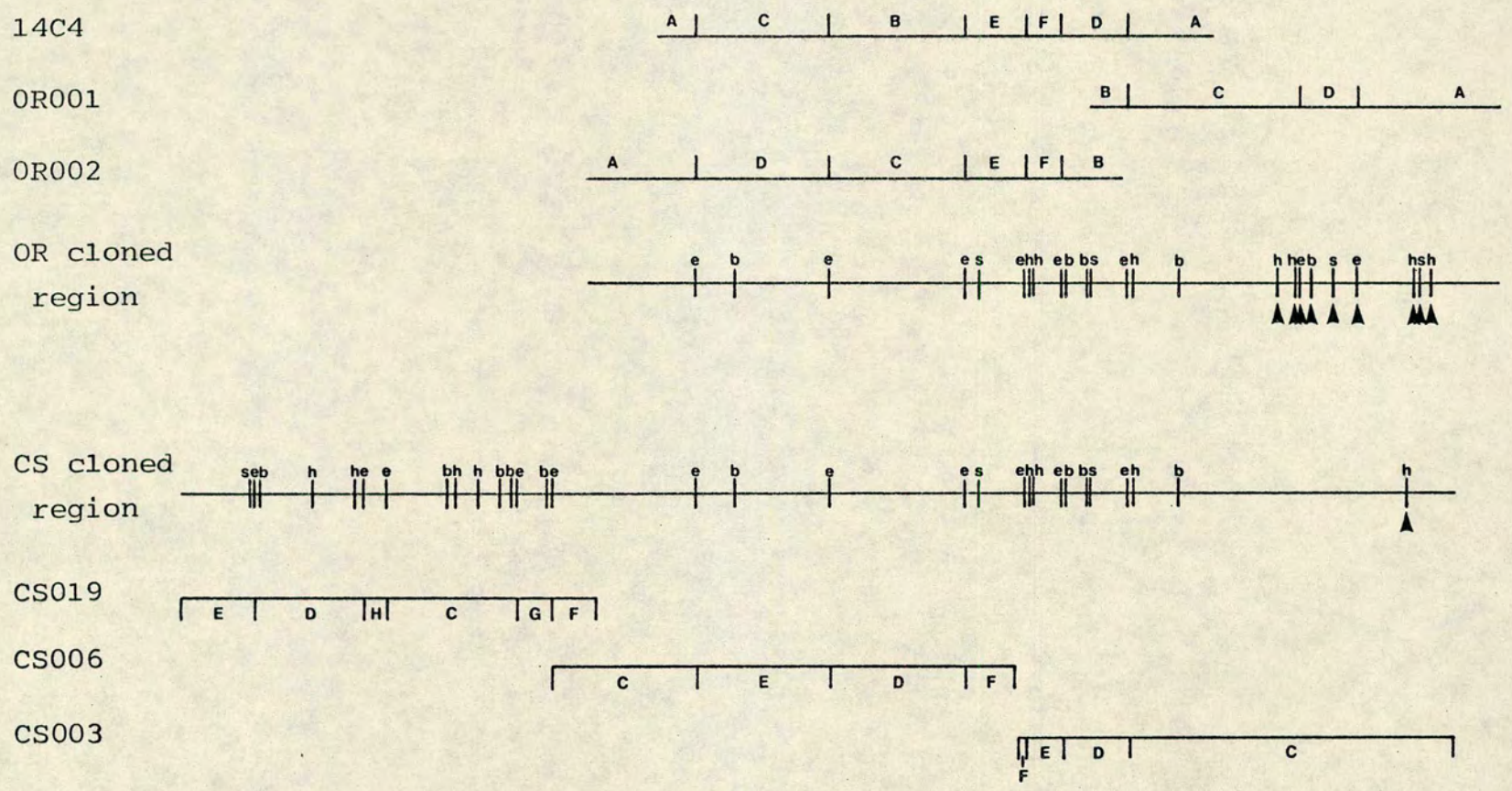

pDm2
A C B D A
OR006
A C E B
OR cloned region
e eh e e s b h h
CS cloned region
h bh s sbh h s h e eh e e s h he s hs e h s s
CS009
D G F H C E
CS007
B G F H I J K D E A
CS005
E C D

The inserts of 14C4 and OR002 were both originally derived from Oregon R embryo DNA. The distance between the rightmost Sma 1 site and the junction between the insert and vector in OR002 is 1.5kb. Therefore the Eco R1 and Hind 111 sites found at the right end of 14C4 would be expected to occur in OR002. This is not the case. However, the distance from the Sma 1 to the junction between insert and vector in OR002 was determined from the length of the Sma 1 fragment lying across the junction; this was measured as 7.2kb. Given the 10% variation between measured sizes determined on different gels, the junction between insert and vector may lie closer to the Sma 1 site than shown in figure 4.30, and the Eco R1 and Hind 111 sites in question lie outside the insert of OR002. This will be assumed to be the case. The additional Hind 111 site in 14C4 which generates the 0.15kb fragment is presumed to exist in OR002 - see the legend to figure 4.30 for the justification for this. Thus the apparent discrepancies between the restriction maps of 14C4 and OR002 can be accounted for, and the recombinants aligned as shown in figure 4.36.

The three recombinants carrying Canton S embryo DNA sequences cannot be positioned on the basis of their restriction maps alone. However, CS003 and CS006 align perfectly with both 14C4 and OR002, and are positioned as shown in figure 4.36 on that basis. If CS003 and CS006 overlap at all, it must be by less than 0.05kb. The restriction map of CS019 does not permit improvement on the region of overlap given in section 4.1. However it should be noted that the leftmost Eco R1 site within the insert of CS019 lies very close to the junction between insert and vector in CS006. It is possible that this site is actually present in CS006, but lies so close to the junction between insert and vector that it is below the limit of resolution of the map.

Figure 4.36
Restriction maps of the regions of the genome from Drosophila melanogaster strains Canton S (CS) and Oregon R (OR) represented by the 14C4 group of recombinants (14C4, CS003, CS006, CS019, OR001, and OR002). Sites for the enzymes Bam H1 (b), Eco R1 (e), Hind 111 (h), and Sma 1 (s) are shown. The regions represented by the individual recombinants are indicated and the Eco R1 fragments of each are labelled alphabetically in decreasing order of size. Arrowheads mark sites present in the cloned region from one strain but not the other.
Scale 1cm = 2kb.

Figure 4.36


14C4
A C B E F D A
0R001
B C D A
0R002
A D C E F B
OR cloned region
e b e e s ehh eb bs eh b h heb s e hsh
CS cloned region
seb h he e bh h bbe be e b e e s ehh eb bs eh b h
CS019
E D H C G F
CS006
C E D F
CS003
E D C F

OR001.

The restriction map of OR001 shares a Bam H1, Eco R1, and Hind 111 site with 14C4. These sites are also found in CS003, but beyond this point the maps of OR001 and CS003 are completely different. The insert of OR001 is therefore believed to extend beyond 14C4 in the same direction as CS003, but to represent an alternative arrangement of sequence at this position. The results set out in section 4.1 are in complete agreement with this conclusion. It is not possible to say whether this is a strain specific difference, or due to heterogeneity within the strains. A double insert in either OR001 or CS003, or a repeated sequence at the right end of 14C4 could also be responsible (see sections 4.4 and 4.5).

Hence OR001 is assigned to the position shown in figure 4.36, and will henceforth be considered a member of the 14C4 group.

CS016 and CS020.

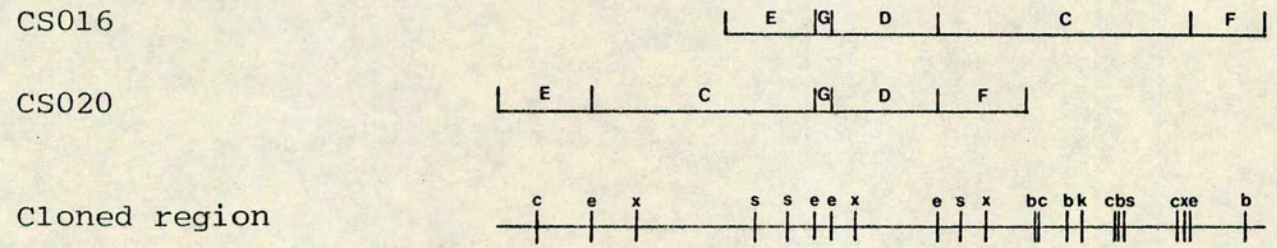
The restriction maps of CS016 and CS020 align exactly with each other as shown in figure 4.37. They bear no relationship with the maps of any other recombinants.

OR028.

The restriction map of OR028 cannot be aligned with that of any other recombinant.

Figure 4.37
Restriction map of the region of the genome from Drosophila melanogaster strain Canton S represented by CS016 and CS020. Sites for the enzymes Bam H1 (b), Eco R1 (e), Kpn 1 (k), Sac 1 (c), Sma 1 (s) and Xho 1 (x) are shown. The regions represented by the individual recombinants are indicated and the Eco R1 fragments of each labelled alphabetically in decreasing order of size.

Figure 4.37


CS016
E
G
D
C
F
CS020
E
C
G
D
F
Cloned region
c
e
x
s
s
e
e
x
e
s
x
bc
bk
cbs
cxe
b

4.3 Cross-hybridisation of recombinants.

DNA from recombinants pDm2, 14C4, CS003, CS005, CS006 CS007, CS009, CS016, CS019, CS020, CS021, OR001, OR002, OR006, and OR028 was digested with Eco R1 (except in the cases of CS007 and OR006 where Hind 111 was used owing to the better distribution of sites), electrophoresed through 0.7% agarose gels and transferred to nitrocellulose filters. Fifteen such filters were prepared. Each filter was probed with nick translated DNA from one of the above recombinants so that all possible pairwise hybridisations were carried out. The results are summarised in table 4.4. Selected tracks of particular interest are shown in figures 4.38, 4.39, and 4.40.

Table 4.4

Key.

s - Strong hybridisation, indicating complete or almost complete homology. Expected from restriction mapping data.

w - Weak hybridisation, indicating partial homology. Expected from restriction mapping data.

w^* - Weak hybridisation, indicating partial homology. Not expected from restriction mapping data.

0 - No detectable homology.

? - Fragment does not appear on the autoradiograph, homology cannot be determined.

Notes.

(a) Not reproducible.

(b) Extremely weak hybridisation.

(c) Vector fragments.

(d) These tracks contain many additional faint bands which are believed to be due to Hind 11 contamination of the Hind 111 used. The extra bands make it difficult to detect weakly hybridising bands; the sensitivity is somewhat reduced.

(e) It is difficult to determine which of these two bands is being labelled. However the hybridisation is largely, if not entirely, due to homology between the vector sequences, and is therefore of little importance.

(f) Hybridisation stronger than expected by comparison with neighbouring bands.

Table 4.4

Cross-hybridisation of pDm2, 14C4, CS003, CS005, CS006, CS007, CS009, CS016, CS019, CS020, CS021, OR001, OR002, OR006, and OR028.

Filter bound fragment		Probe DNA														
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
pDm2	A	s	w*a	w*a	w	0	w	w	0	0	0	0	0	0	w	0
	B	s	0	w*a	0	0	w	s	0	0	0	0	0	0	s	0
	C	s	0	0	0	0	s	s	0	0	0	0	0	0	s	0
	D	s	0	0	0	0	0	s	0	0	0	0	0	0	s	0
	E	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
14C4	A	0	s	w	0	w	0	0	0	0	0	0	w	w	0	0
	B	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	C	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	D	0	s	s	0	0	0	0	0	0	0	0	w	s	0	0
	E	0	s	w	0	s	0	0	w*	0	w*	0	0	s	0	w*
	F	0	s	s	0	w*	0	0	w*b	0	w*	0	0	s	0	w*
CS003	A[c]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[c]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	w	s	0	0	w*b	0	0	0	0	w*b	w	0	0	0
	D	0	s	s	0	0	0	0	0	0	0	0	w	w	0	0
	E	0	s	s	0	w*	0	0	w*b	0	w*	0	0	s	0	w*
	F	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Table 4.4 continued

Filter bound fragment		Probe DNA														
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
CS005	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	0	0	s	0	0	w	0	0	0	0	0	0	w	0
	D	0	0	0	s	0	0	0	0	0	0	0	0	0	0	0
	E	w	0	0	s	0	0	s	0	0	0	0	0	0	s	0
CS006	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	w	0	0	s	0	0	0	w	0	0	0	w	0	0
	D	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	E	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	F	0	s	w*	0	s	0	0	w*	0	w*	0	0	s	0	w*

Table 4.4 continued

Filter bound fragment		Probe DNA														
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
CS007[d]	A	w	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B	0	0	s	s	s	s	s	s	s	s	s	s[e]	s[e]	s[e]	s[e]
	C[c]	0	0	s	s	s	s	s	s	s	s	s	0[e]	0[e]	0[e]	0[e]
	D	0	0	0	0	0	s	w	0	0	0	0	0	0	0	0
	E	w	0	0	0	0	s	s	0	0	0	0	0	0	w	0
	F	0	0	0	0	0	s	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	s	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	s	0	0	0	0	0	0	0	0	0
I,J,K		?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
CS009	A[c]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[c]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	s	0	0	w	0	0	s	0	0	0	0	0	0	s	0
	D	w	0	0	0	0	s	s	0	0	0	0	0	0	w	0
	E	0	0	0	s	0	0	s	0	0	0	0	0	0	s	0
	F	s	0	0	0	0	w	s	0	0	0	0	0	0	s	0
	G	s	0	0	0	0	s	s	0	0	0	0	0	0	s	0
	H	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Table 4.4 continued

Filter bound fragment		Probe DNA														
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
CS016	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	0	0	0	0	0	0	s	0	w	0	0	0	0	0
	D	0	w*	w*	0	w*	0	0	s	0	s	0	0	w*	0	w*
	E	0	0	0	0	0	0	0	s	0	s	0	0	0	0	0
	F	0	0	0	0	0	0	0	s	0	0	0	0	0	0	0
	G	0	0	0	0	0	0	0	s	0	s	0	0	0	0	0
CS019	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	0	0	0	0	0	0	0	s	0	0	0	0	0	0
	D	0	0	0	0	0	0	0	0	s	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0	s	0	0	0	0	0	0
	F	0	0	0	0	s	0	0	0	s	0	0	0	w	0	0
	G	0	0	0	0	0	0	0	0	s	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	s	0	0	0	0	0	0

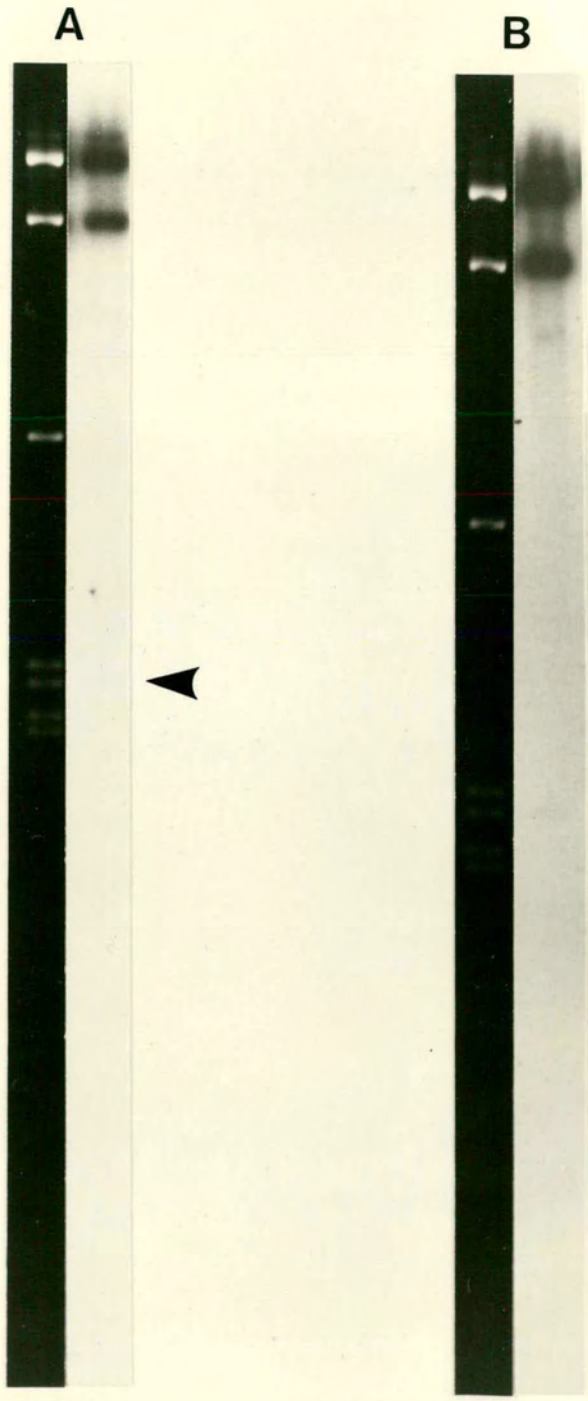
Table 4.4 continued

Filter bound fragment		Probe DNA														
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
CS020	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	w*	w*	0	w*	0	0	w	0	s	0	0	w*	0	w*
	D	0	w*	w*	0	w*	0	0	s	0	s	0	0	w*	0	w*
	E	0	0	0	0	0	0	0	0	0	s	0	0	0	0	0
	F	0	0	0	0	0	0	0	s[f]	0	s	0	0	0	0	0
	G	0	0	0	0	0	0	0	s	0	s	0	0	0	0	0
CS021	A[C]	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B[C]	0	0	s	s	s	s	s	s	s	s	s	w	w	w	w
	C	0	0	0	0	0	0	0	0	0	0	s	0	0	0	0
	D	0	0	0	0	0	0	0	0	0	0	s	0	0	0	0
	E	0	0	w*	0	w*	0	0	0	0	0	s	0	0	0	0
	F	0	0	0	0	0	0	0	0	0	0	s	0	0	0	0
	G	0	0	0	0	0	0	0	0	0	0	s	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	s	0	0	0	0
OR001	A	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B	0	w	w	w	w	w	w	w	w	w	w	s	s	s	s
	C	0	w	w	0	0	0	0	0	0	0	0	s	0	0	0
	D	0	0	0	0	0	0	0	0	0	0	0	s	0	0	0

Table 4.4 continued

Filter bound fragment		Probe DNA													
		pDm2	14C4	CS003	CS005	CS006	CS007	CS009	CS016	CS019	CS020	CS021	OR001	OR002	OR006	OR028
OR002	A	0	w	s	s	s	s	s	s	s	s	s	s	s	s	s
	B	0	w	w	w	w	w	w	w	w	w	w	w	s	s	s
	C	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	D	0	s	0	0	s	0	0	0	0	0	0	0	s	0	0
	E	0	s	w	0	s	0	0	w*	0	w*	0	0	s	0	w*
	F	0	s	s	0	w*	0	0	w*[b]	0	w*	0	0	s	0	w*
OR006[d]	A	w	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B	0	0	w	w	w	w	w	w	w	w	w	w	w	s	w
	C	s	0	0	w	0	w	s	0	0	0	0	0	0	s	0
	D[c]	0	0	w	w	w	w	w	w	w	w	w	s	s	s	s
	E	s	0	0	s	0	0	s	0	0	0	0	0	0	s	0
OR028	A	0	0	s	s	s	s	s	s	s	s	s	s	s	s	s
	B	0	0	w	w	w	w	w	w	w	w	w	w	w	w	s
	C	0	w*	w*	0	w*	0	0	w*	0	w*	0	0	w*	0	s
	D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s
	E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s
	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s
	G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s

Figure 4.38
DNA from recombinant CS021 digested with Eco R1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated CS003 DNA (A) or CS007 DNA (B): gel photographs (left) and autoradiographs (right). Note the hybridisation to the E fragment (arrowheads).

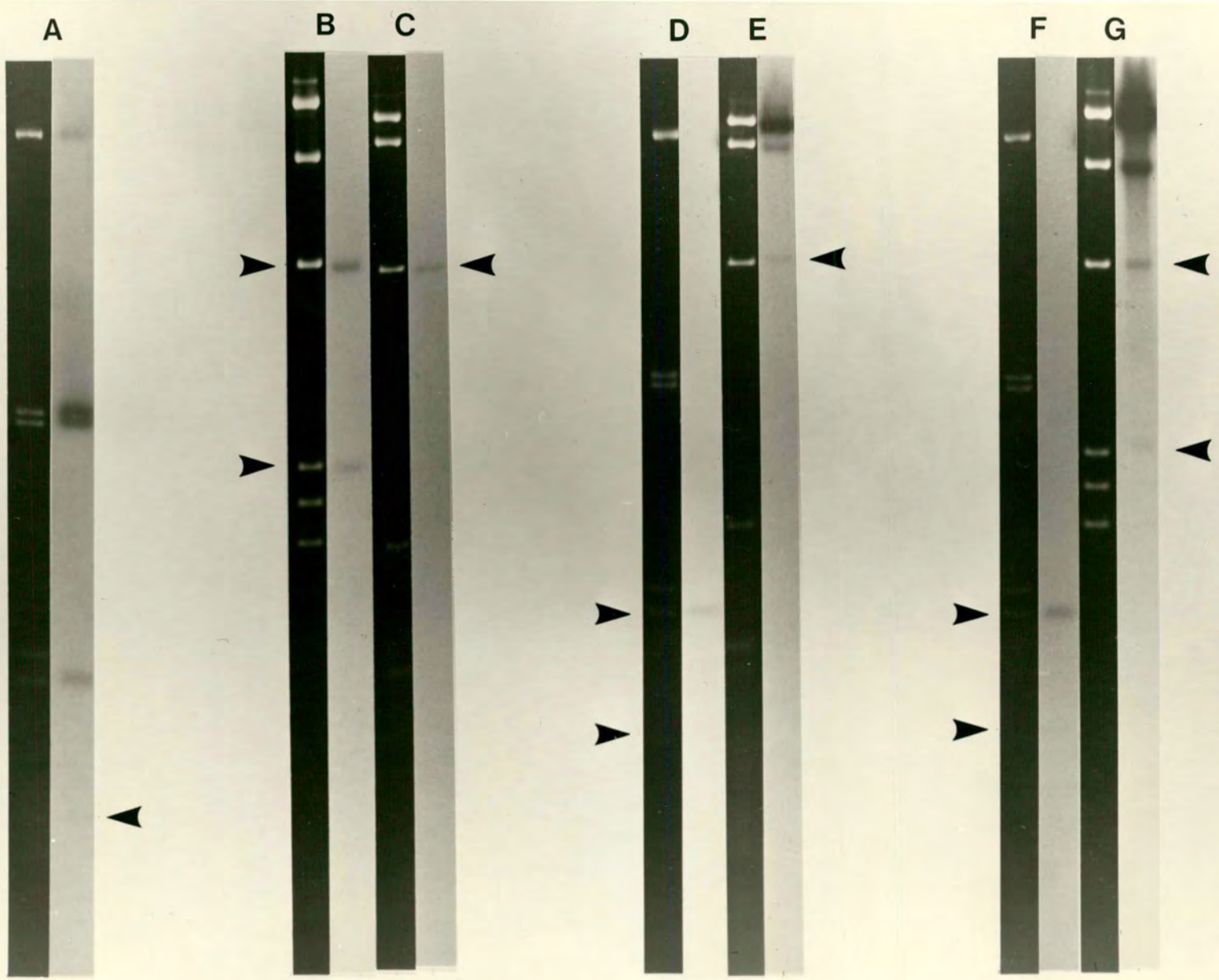


4·38

Figure 4.39

DNA from recombinants 14C4, CS020, and OR028 digested with Eco R1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated 14C4, CS006, CS020 or OR028 DNA: gel photographs (left) and autoradiographs (right). Arrowheads indicate hybridisation to bands not expected to show homology on the basis of the regions of overlap shown in figures 4.35, 4.36 and 4.37.

Track	Probe DNA	Filter bound DNA	Fragments showing unexpected hybridisation
A	CS006	14C4	F
B	14C4	CS020	C, D
C	14C4	OR028	C
D	CS020	14C4	E, F
E	CS020	OR028	C
F	OR028	14C4	E, F
G	OR028	CS020	C, D

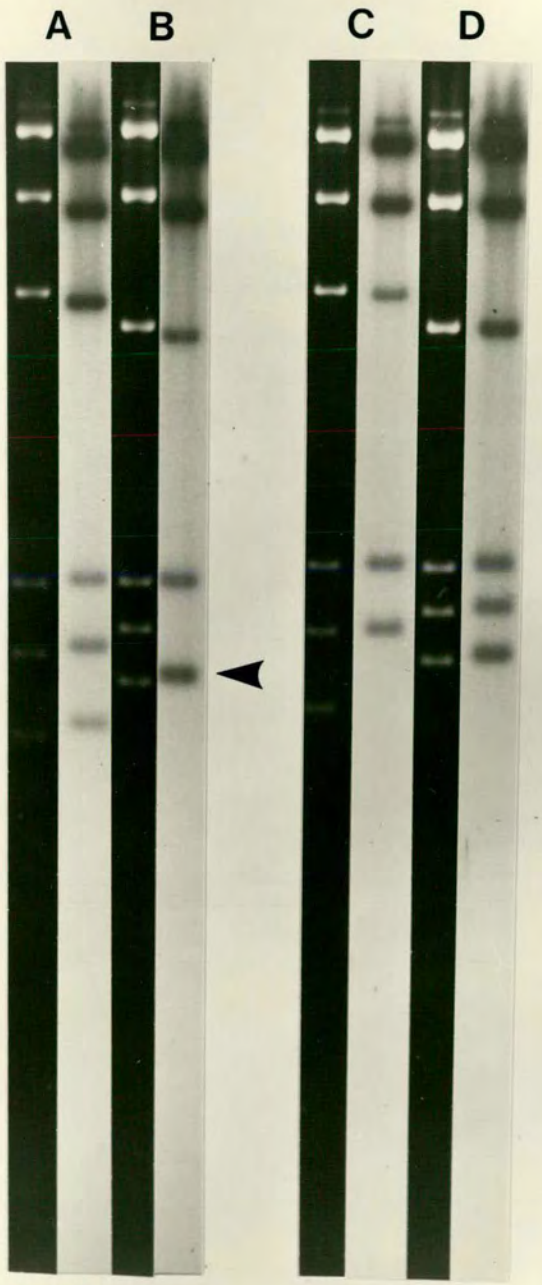


4·39

Figure 4.40

DNA from recombinants CS016 and CS020 digested with Eco R1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated CS016 or CS020 DNA: gel photographs (left) and autoradiographs (right). Note the intensity of hybridisation to the F fragment of CS020 with the CS016 DNA probe (arrowhead).

Track	Probe DNA	Filter bound DNA
A	CS016	CS016
B	CS016	CS020
C	CS020	CS016
D	CS020	CS020



4·40

Discussion.

From the positions of overlap shown for the various recombinants in figures 4.35, 4.36, and 4.37, a particular pattern of cross-hybridisation would be predicted. Experimentally, all fragments which are expected to be labelled with a particular probe DNA are so labelled, and, with one exception, to the expected level of intensity (see table 4.4). The exception is the F fragment of CS020, which, when probed with nick translated CS016 DNA, labels more intensely (by comparison with adjacent bands) than would be expected were it to be completely homologous with the probe. A possible explanation for this will be given below.

In addition, several fragments which were not expected to hybridise with particular probes in fact did. These anomalous cross-hybridisations can be broken down into two groups: (a) those involving CS003 (C fragment), CS007, and CS021 (E fragment), and (b) those involving the E and F fragments of 14C4, their equivalents in CS003, CS006, and OR002, CS016 (D fragment), CS020 (C and D fragments), and OR028 (C fragment). The hybridisation of the A and B fragments of pDm2 with 14C4 and CS003 DNA probes was not reproducible, and will be taken to be an artifact.

Group (a).

The E fragment of CS021 is labelled when probed with nick translated CS003 or CS007 DNA (figure 4.38). The C fragment of CS003 is labelled much less intensely by CS007 or CS021 DNA probes (hybridisation too weak to appear on reproduction; results not shown). No fragments of CS007, except those containing vector DNA, are labelled detectably by the other members of the group, although clearly homology must be present. There are several possible explanations for this: the homology may lie in the insert sequences of fragments A and/or C, and be lost in the vector hybridisation; the homology may be on fragments I, J, and/or K, which are too small to appear on the autoradiographs; the hybridising fragment may be

lost in the background of Hind 11 partials.

These results indicate that there must be repeated sequences present on these three recombinants. From the intensity of hybridisation the repeat must be either very short (less than 0.5kb) or very poorly conserved. The results are most simply explained by suggesting that there is a single repeated sequence, a copy of which is present on all three recombinants. It is equally possible that several different repeated sequences could be involved.

The C fragment of CS003 is labelled much less intensely by the CS021 DNA probe than the E fragment of CS021 in the reciprocal experiment. The most likely explanation for this apparent discrepancy is inefficient transfer of the CS003 fragment. This fragment is considerably larger than the E fragment of CS021 (9.3 as against circa 2.3kb). Large fragments are transferred less efficiently than small ones. Thus there would be fewer copies of the C fragment available for hybridisation on the filter, and the intensity of labelling would be reduced. A similar result might be obtained were there to be more copies of the repeated sequence on one recombinant than on the other. This is an extremely complex situation and difficult to analyse. Suffice to say that this possibility must be borne in mind when attempting to draw conclusions from the relative intensities of hybridisation in the absence of other data. Note also that the C fragment of CS003 is labelled less intensely than would be expected by comparison with neighbouring fragments when probed with CS003 DNA. This result is not understood.

It is appropriate now to reconsider the isolation of CS021. This recombinant was isolated in the second screening experiment using a mixed probe of 14C4, CS005f, and CS006f DNAs (see section 4.1), yet it shows no homology with any fragment of these recombinants. In the initial plaque hybridisation CS021 gave a moderately strong signal, about equal to that of CS019. This suggested a region of homology of about 1kb. During plaque purification this homology was not apparent;

CS021 was retained more because it could not be said to be definitely negative than because it was positive. One is therefore forced to conclude that the initial signal was either due to another recombinant, now lost, or that it was an artifact, and that CS021 was picked up by chance. The likelihood that a recombinant picked up at random will contain a sequence homologous to that carried by another recombinant depends on the frequency of repetition and distribution of that sequence. If just one repeated sequence is responsible for the cross-hybridisation observed, than, were the sequence to be repeated one thousand times in the genome, the probability of isolating by chance a recombinant carrying a copy of this sequence is very approximately one in ten.

Group (b).

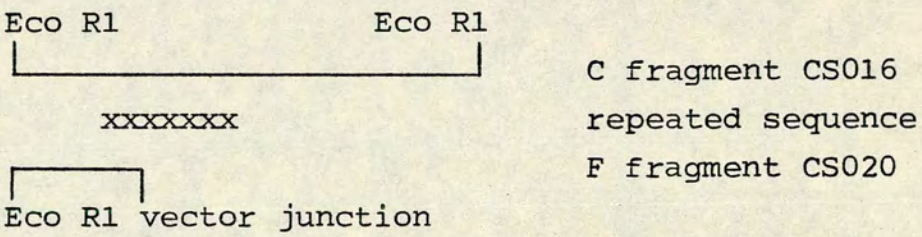
Taking the recombinants 14C4, CS003, CS006, and OR002 first, it is clear that there is a sequence within the F fragment of CS006 (and hence within the corresponding fragments of 14C4 and OR002) which is repeated within the F fragment of 14C4 (and corresponding fragments of CS003 and OR002). With the strictures mentioned above, from the relative intensities of hybridisation (see figure 4.39), the repeated sequence (which may well extend beyond the end of CS006 into the F fragment of CS003) must be less than 0.3kb long, assuming the repeat to be perfect. Obviously, the more imperfect the repeat, the greater the maximum size.

It will be recalled from section 1 that there is a cluster of tRNA genes on the E and F fragments of 14C4 (and corresponding fragments of CS003, CS006, and OR002) (P. Schedl, personal communication, see also section 4.6). This cluster consists of at least two arginine and two asparagine tRNA genes. It seems quite likely, although of course not certain, that these genes could form part, if not all, of the repeated sequence. The extent to which different tRNA genes cross-hybridise under the conditions used is not known.

Turning to the other members of this group, any fragment having homology with any one member of the group shows homology with all other members of the group (excepting fragments which are expected to cross-hybridise by virtue of regions of overlap shown in figures 4.36 and 4.37), although the relative intensities vary slightly (see figure 4.39). No simple model can be proposed which satisfactorily accounts for these results in detail. Multiple copies, possibly of varying fidelity, of at least two, and more probably three or more different repeated sequences are likely to be involved. Some or all of these sequences could be tRNA genes.

Finally there is the question of the excessive labelling of the F fragment of CS020 with the CS016 DNA probe. The F fragment is clearly labelled more heavily than the D fragment (see figure 4.40), although they would both be expected to be completely homologous with the probe from the restriction map shown in figure 4.37. It does not appear that the D fragment is under-labelled; by comparison with the D, E, and F fragments of CS016 it seems exactly right. The most likely explanation is that on fragment F of CS020 there is a sequence which is present in several copies on CS016, and that the conditions of hybridisation were such that the reaction did not go to completion. The sequence on fragment F of CS020, being present in several copies in the probe DNA, would be at a higher concentration than the unique sequences, and would thus drive its hybridisation further towards completion, resulting in more intense labelling of the F fragment. As the F fragment of CS020 is not labelled excessively when probed with CS020 DNA (the labelling may be slightly more intense than expected, but certainly not to the extent found with the CS016 DNA probe), the extra copies of the proposed repeated sequence must lie outside the region common to CS016 and CS020. This places them within the C and/or F fragments of CS016. The F fragment of CS016 is not labelled excessively when probed with CS016

DNA so the repeated sequence must lie on the C fragment. There are no suitable fragments by which the intensity of hybridisation of this fragment when probed with CS016 DNA may be judged. Being a larger fragment, any effect would be less noticeable in any case. Note that the C fragment of CS016 and the F fragment of CS020 are co-linear (see figure 4.37); one might envisage a tandemly repeated sequence as illustrated below.



4.4 Identification of Repetitive Sequences.

Total genomic DNA from Oregon R or Canton S strain embryos was digested with selected restriction enzymes and electrophoresed through 0.7% agarose gels. In one case, as a control against partial digestion or contamination of the restriction enzyme (see section 1), two tracks of embryo DNA were used, one being digested with that amount of enzyme supposed to give complete digestion, the other with twice that amount. In one series of experiments unlabelled samples of the recombinant DNAs to be used as probes were digested with the same restriction enzymes and run alongside the appropriate embryo DNA tracks, the amount of DNA used being equivalent to that of single copy sequences in the embryo DNA tracks. The DNA was transferred to nitrocellulose filters and probed with nick translated pDm2, 14C4, CS003, CS005, CS006, CS007, or OR001 DNA. Under the conditions used (5ug embryo DNA, probes of specific activity of at least 1.5×10^7 cpm per ug), fragments having homologies greater than 0.2kb would be expected to be detected, assuming that the hybridisation reaches 10% of completion. The results are shown in figures 4.41 to 4.44, and summarised in table 4.5.

These experiments give an indication of the restriction sites within the genome adjacent to which the sequences used as probe are found. For a unique sequence the expected number of fragments on the autoradiograph is equal to the number of segments of genomic sequence produced by digestion of the probe with the given restriction enzyme. Where the inserts of the library were not generated by digestion with this same enzyme, as is the case here, two new fragments will appear in place of the two (or one, depending on the restriction enzyme and vector) fragments carrying the terminal genomic segments of the insert; these indicate

the distance to the adjacent restriction sites (see figure 1.2). The expected sizes of these fragments given in table 4.5 are based on the restriction maps shown in figures 4.35 and 4.36. Three assumptions are made in arriving at these predictions: that the sequences used as probe are unique; that they represent the only arrangement of sequences in the population of embryos from which the genomic DNA was obtained; that all restriction sites are invariant, that is, the sequence is highly conserved. The experiments are designed to test these assumptions.

The results observed may differ from those expected in (1) the number of fragments, (2) their sizes, and (3) the intensities with which they hybridise. Each of these possibilities will now be considered in turn.

(1) Discrepancies in the number of bands have already been discussed in section 1. To reiterate: if there are more new bands than expected then either (a) the probe contains a repetitive sequence, (b) alternative arrangements of the sequence to that found on the probe occur within the population of embryos from which the genomic DNA was obtained, (c) the sequence is not highly conserved, and additional restriction sites are present within the sequence in some members of the population, (d) the probe contains a double insert, (e) the genomic DNA was only partially digested by the restriction enzyme, or (f) the restriction enzyme used contained a contaminating activity. Alternative arrangements of sequence may be due to heterogeneity within the population or be developmentally specific. If the number of bands is fewer than expected, a restriction site has been lost indicating lack of sequence conservation.

(2) If restriction fragments of the expected sizes do not appear on the autoradiograph, then the pattern of restriction sites observed on the probe (or adjacent) recombinant must at best be infrequent within the population. The maximum frequency with which a particular arrangement could occur and still not be detected is

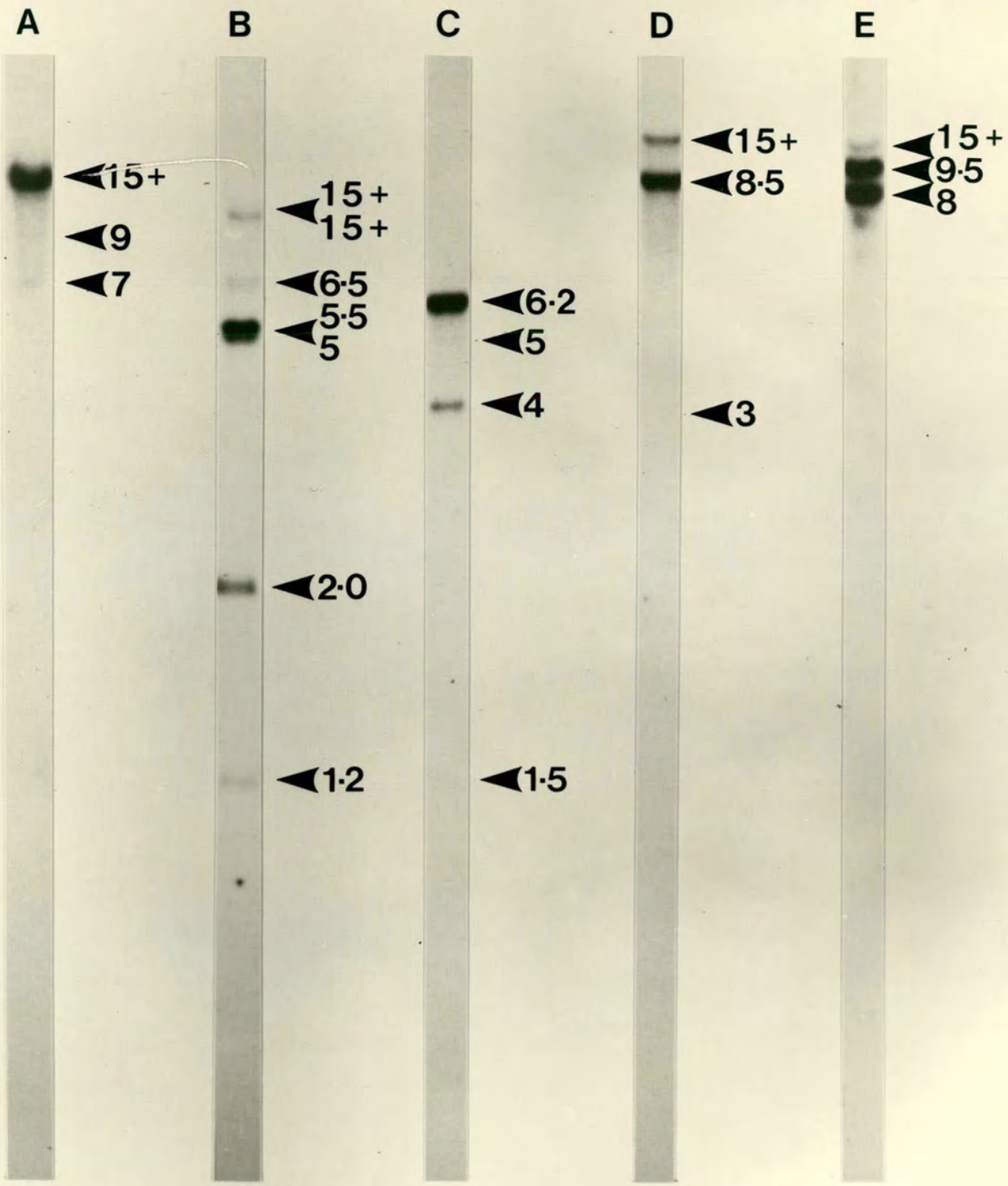
dependent on the length of homology carried by the missing fragment. The pattern of restriction sites may vary by virtue of alternative arrangements of sequence, the breakpoint occurring within the missing fragment, or by lack of sequence conservation leading to gain or loss of restriction sites. At worst, a double insert in the probe (or adjacent) recombinant might be responsible,the breakpoint again lying within the missing fragment.

(3) The intensity of hybridisation gives an indication of the degree of repetition. Unique fragments will show a certain level of hybridisation dependent on the length of homology, assuming the sequence to be reasonably conserved. If the probe carries a repeated sequence, the additional copies of the repeat will also hybridise, the intensity depending on the length and fidelity of the repeat. Hence the overall intensity in the track will be greater than anticipated. Moreover, if a particular fragment is labelled more heavily than expected from its length, it can be said that the sequences on this fragment are often repeated intact. There is an exception to this; where a sequence is repeated within the probe, fragments carrying copies of the repeat may be labelled with extra density - see discussion of the labelling of the F fragment of CS020 with CS016 DNA in section 4.3. If a fragment is labelled less heavily than expected from its overlap with the probe sequences, this indicates that it is not found in all individuals of the strain from which the DNA was derived. This in turn means that there must be some variation in the strain within the region represented by that fragment.

Figure 4.41

Total genomic DNA from Oregon R strain embryos digested with Bam H1, Eco R1, Hind 111, Pst 1, or Sma 1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated pDm2 DNA: autoradiographs. Arrowheads give the approximate sizes in kb of the fragments observed.

Track	Enzyme	Fragments contained within pDm2
A	Bam H1	none
B	Eco R1	2.0, 1.2
C	Hind 111	6.2
D	Pst 1	none
E	Sma 1	none

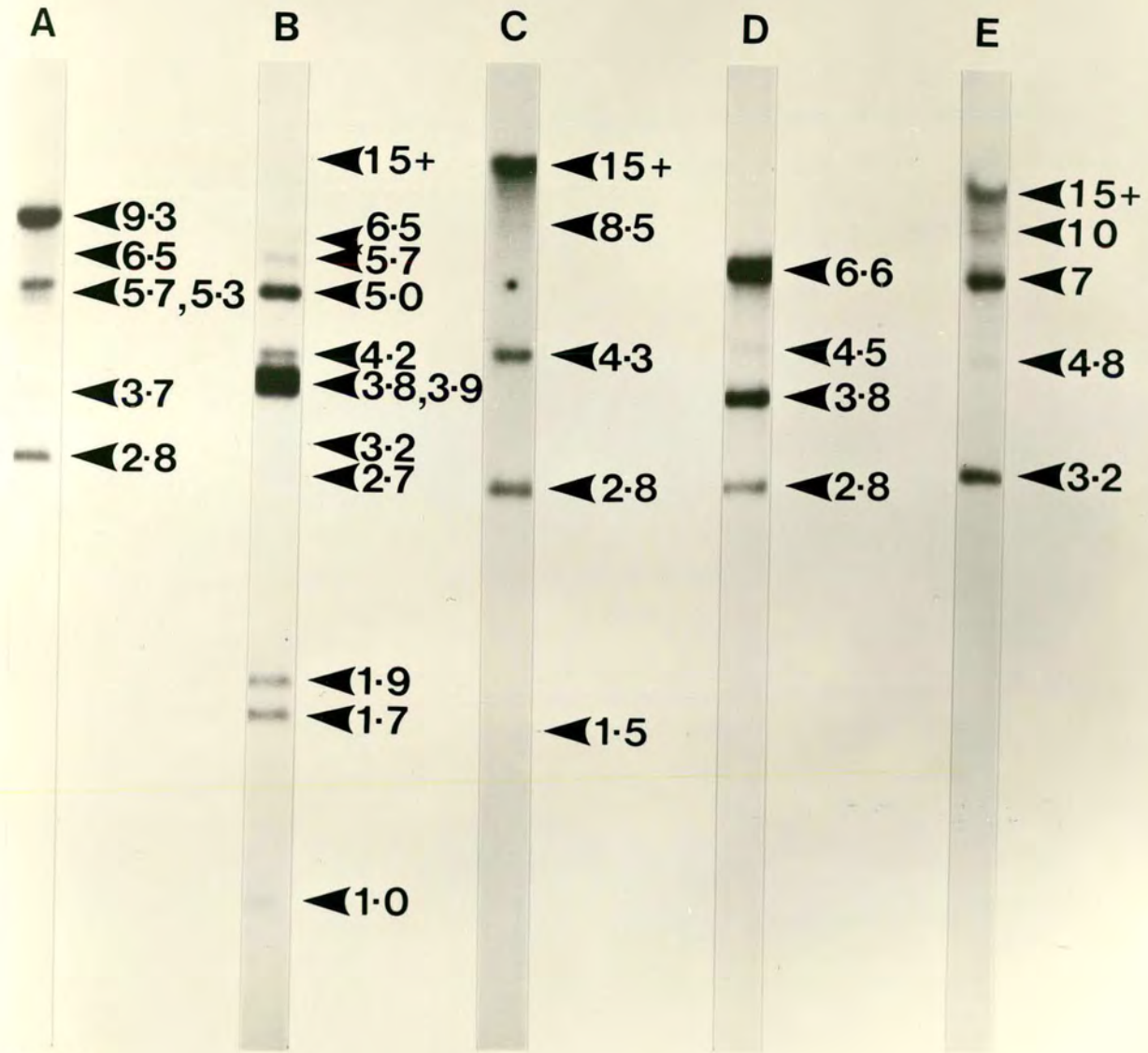


4·41

Figure 4.42

Total genomic DNA from Oregon R strain embryos digested with Bam H1, Eco R1, Hind 111, Pst 1, or Sma 1, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated 14C4 DNA: autoradiographs. Arrowheads give approximate sizes in kb of the fragments observed.

Track	Enzyme	Fragments contained within 14C4
A	Bam H1	9.3, 2.8
B	Eco R1	3.9, 3.8, 1.9, 1.7, 1.0
C	Hind 111	2.8
D	Pst 1	6.6, 3.8
E	Sma 1	3.2

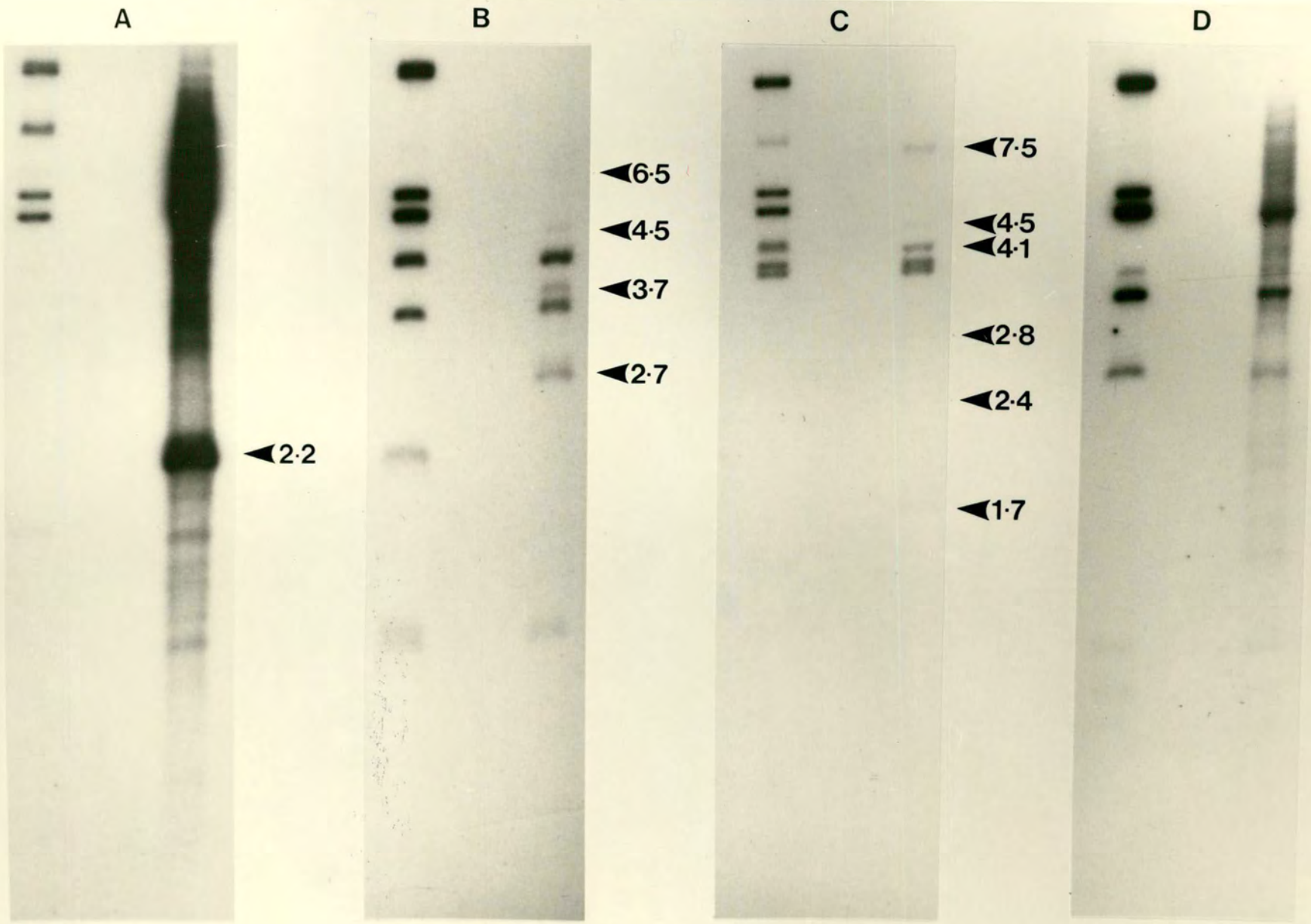

A
B
C
D
E
9·3
6·5
5·7,5·3
3·7
2·8
15+
6·5
5·7
5·0
4·2
3·8,3·9
3·2
2·7
1·9
1·7
1·0
15+
8·5
4·3
2·8
1·5
6·6
4·5
3·8
2·8
15+
10
7
4·8
3·2

4·42

Figure 4.43

DNA from Oregon R strain embryos, and from the recombinants CS003, CS005, CS006 and CS007, digested with Eco R1 and Hind 111, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose, and hybridised with nick translated CS003, CS005, CS006 or CS007 DNA: autoradiographs. The recombinant DNA (indicating the intensity of hybridisation expected for single copy sequences) is at the left, and the genomic DNA at the right in each pair of tracks. Arrowheads give the approximate sizes in kb of fragments not present in the digest of recombinant DNA.

Track	Probe and filter bound DNA
A	CS003
B	CS005
C	CS006
D	CS007

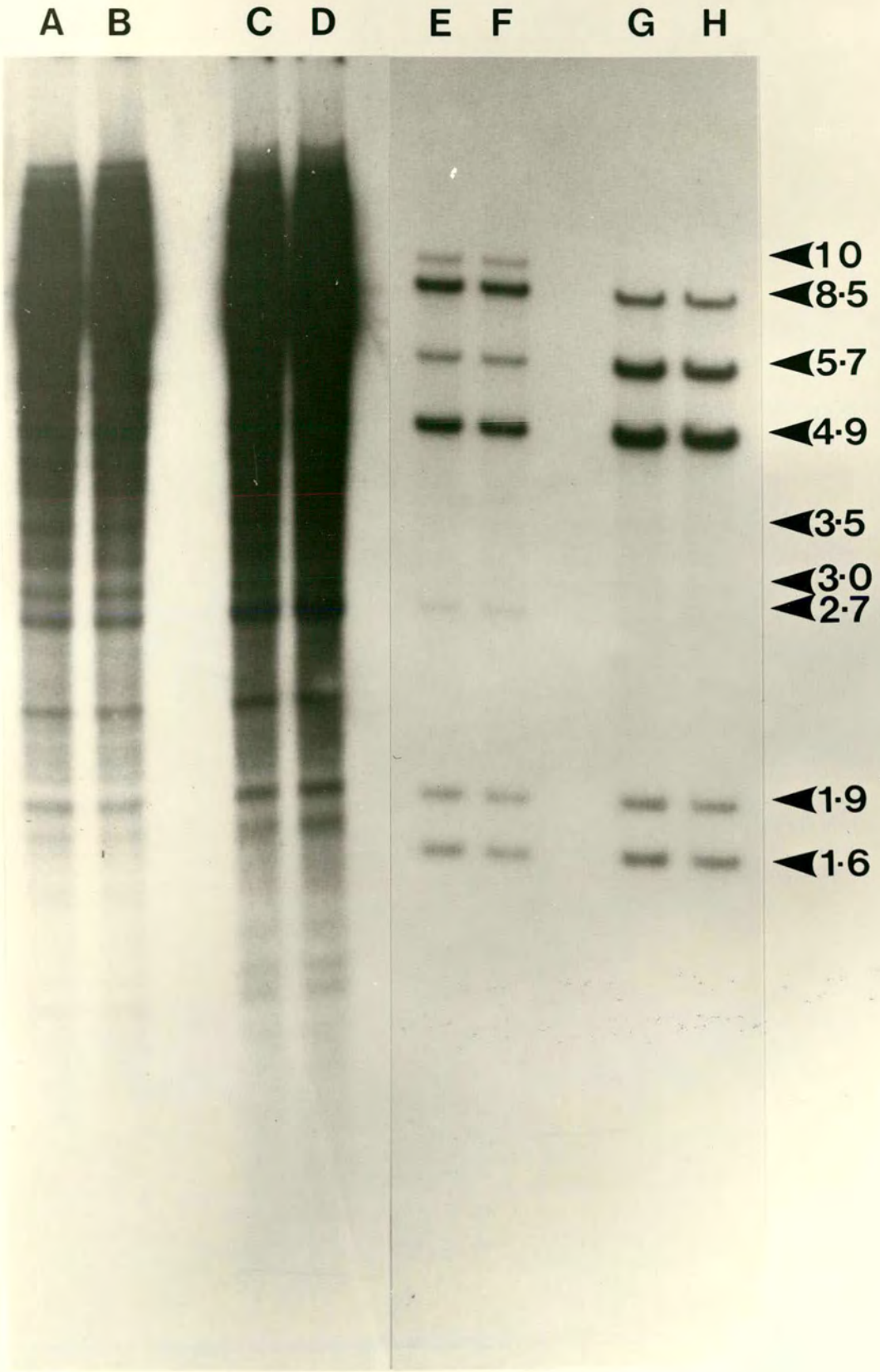


4·43

Figure 4.44

Total genomic DNA from Canton S and Oregon R strain embryos, digested with single and double quantities of Eco R1 (see section 1), electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated CS003 or OR001 DNA: autoradiograph. Arrowheads give the approximate sizes in kb of the fragments observed in tracks E to G. The 4.9 and 1.6kb fragments are derived entirely from sequences within the insert of OR001.

Track	Probe DNA	Filter bound DNA	Amount of Eco R1
A	CS003	Canton S	single
B	CS003	Canton S	double
C	CS003	Oregon R	double
D	CS003	Oregon R	single
E	OR001	Canton S	single
F	OR001	Canton S	double
G	OR001	Oregon R	double
H	OR001	Oregon R	single



4·44

Table 4.5

Notes.

(a) Fragments derived entirely from within the sequence contained on the probe are omitted.

(b) From the autoradiographs shown in figures 4.41 to 4.44. Sizes are estimated by comparison with lambda DNA fragments of known molecular weight run on the same gels; they should be regarded as approximate.

(c) Predicted from the restriction maps of Oregon R DNA shown in figures 4.35 and 4.36. The fragment extending to the left of the probe (as drawn in the figures) is given first.

(d) Predicted from the restriction maps of Canton S DNA shown in figures 4.35 and 4.36. The fragment extending to the left of the probe (as drawn in the figures) is given first.

(e) OR - DNA from embryos of strain Oregon R.
CS - DNA from embryos of strain Canton S.

(f) Restriction enzymes with which the embryo DNA was digested.

(g) This column gives the length of fragment in kb.

(h) This column gives the level of hybridisation.

s - strong
m - moderate
w - weak

(i) This band may be a doublet.

(j) Depending on the exact position of the break in homology between CS003 and OR001.

(k) Due to the break in homology between CS003 and OR001, one or two additional bands are expected.

(l) Lambda molecular weight standards were not included on this gel. Sizes are very approximate.

Table 4.5

Probing of restricted total Drosophila melanogaster embryo DNA with nick translated pDm2, 14C4, CS003, CS005, CS006, CS007, and OR001 DNA.

Probe DNA	Filter bound DNA	Fragment size[a] and level of hybridisation					
		Measured[b]		Expected (OR)[c]		Expected (CS)[d]	
pDm2	OR[e]	15+[g]	s[h]	7.05+[g]	s[h]	33.35+[g]	s[h]
	Bam H1[f]	9	w	5.6	m		
		7	w				
	OR	15+	w	1.4+	w	14.55+	w
	Eco R1	15+	m	7.5+	s	5.35	s
		6.5[i]	w				
		5.5	s				
		5	m				
	OR	5	w	2.6+	m	3.8	m
	Hind 111	4	m	1.4	w	1.4	w
		1.5	w				
	OR	15+	m	0.05+	w	?	?
	Pst 1	8.5	s	8.55+	s	?	?
		3	w				
	OR	15+	w	5.95+	s	9.75	s
	Sma 1	9.5	s	6.7+	s	8.05	s
		8	s				
14C4	OR	6.5	w	4.25+	m	5.45	m
	Bam H1	5.7	m	3.6	w	7.7+	w
		5.3	w				
		3.7	w				

Table 5 continued

Probe DNA	Filter bound DNA	Fragment size and level of hybridisation: Measured		Expected (OR)		Expected (CS)	
14C4	OR	15+	w	3.1+	w	4.2	w
	Eco R1	6.5	w	4.9	m	9.3+	m
		5.7	w				
		5.0	m				
		4.2	w				
		3.2	w				
		2.7	w				
	OR	15+	s	12.8+	s	15.8	s
	Hind 111	8.5	w	4.1	m	7.9	m
		4.3	m				
		1.5	w				
	OR	4.5	w	1.8+	w	?	?
	Pst 1	2.8	m	3.1+	m	?	?
	OR	15+	m	11.2+	s	21.2	s
	Sma 1	10	w	6.95	m	10.4+	m
		7	s				
		4.8	w				
CS003	OR	many		1.7	w	1.7	w
	Eco R1-	bands		4.1	w-s[j]	1.2+	w
	Hind 111			?[k]			
CS005	OR	6.5[i]	w	3.7	w	3.7	w
	Eco R1-	4.5	w	?	?	2.1+	m
	Hind 111	3.7	w				
		2.8	m				

Table 5 continued

Probe DNA	Filter bound DNA	Fragment size and level of hybridisation: Measured		Expected (OR)		Expected (CS)	
CS006	OR	7.5	m	3.1+	s	4.2	s
	Eco R1-	4.5	w	1.7	w	1.7	w
	Hind 111	4.1	s				
		2.8	w				
		2.4	w				
		1.7	w				
CS007	OR	many		?	?	0.4+	w
	Eco R1-	bands		2.0	w	2.0	w
	Hind 111						
CS003	OR	many		1.7	w	1.7	w
	Eco R1	bands		4.9	w-s[j]	9.3+	s
				?[k]			
	CS	many		1.7	w	1.7	w
	Eco R1	bands		4.9	w-s[j]	9.3+	s
				?[k]			
OR001	OR[1]	8.5	m	1.9	w	1.9	w
	Eco R1	5.7	s	4.1+	s	9.3+	w-s[j]
		3.5	w			?[k]	
		3.0	w				
		2.7	w				
		1.9	m				
	CS[1]	10	w	1.9	w	1.9	w
	Eco R1	8.5	s	4.1+	s	9.3+	w-s[j]
		5.7	m			?[k]	
		2.7	w				
		1.9	m				

Figure 4.35

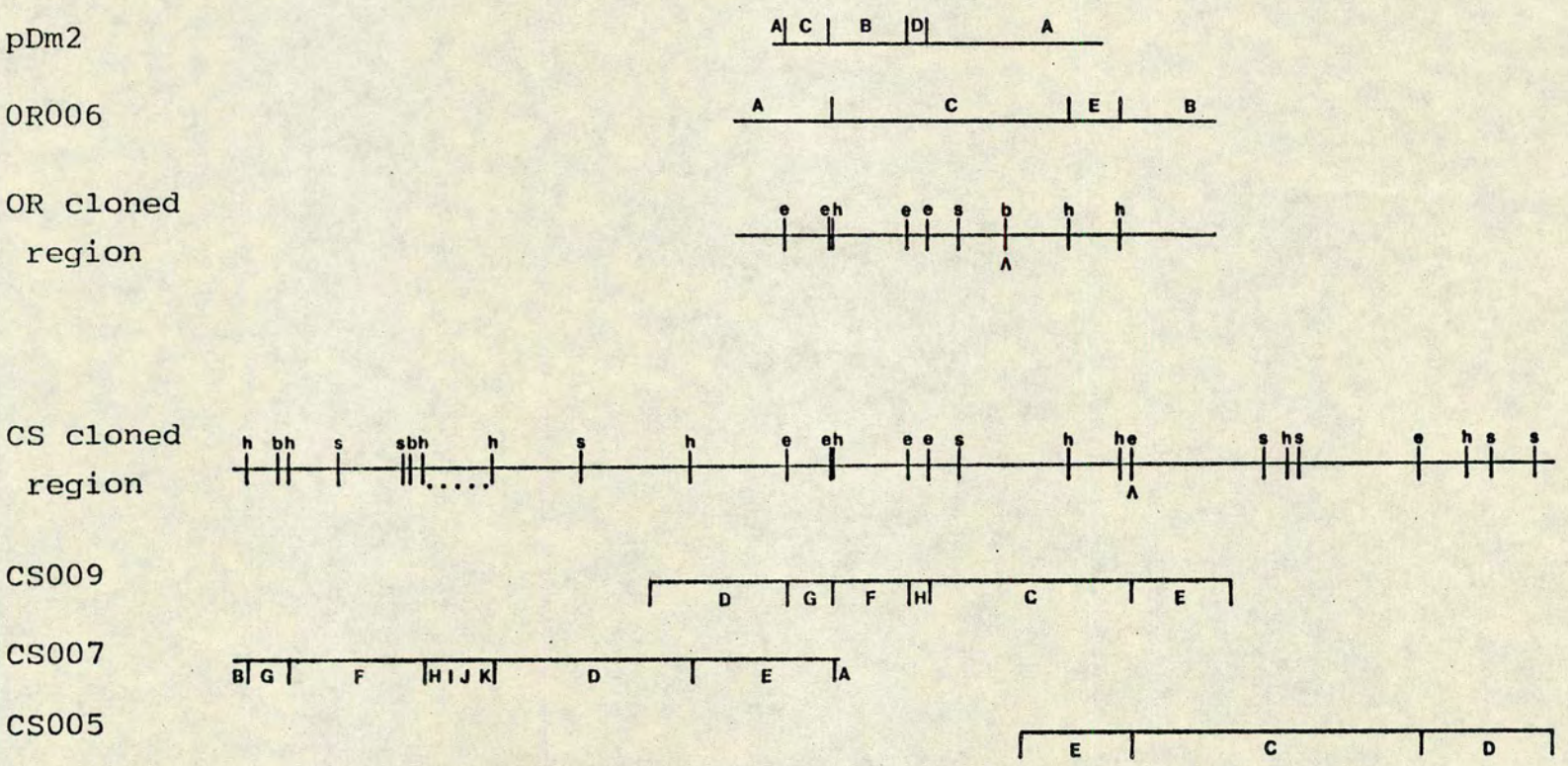
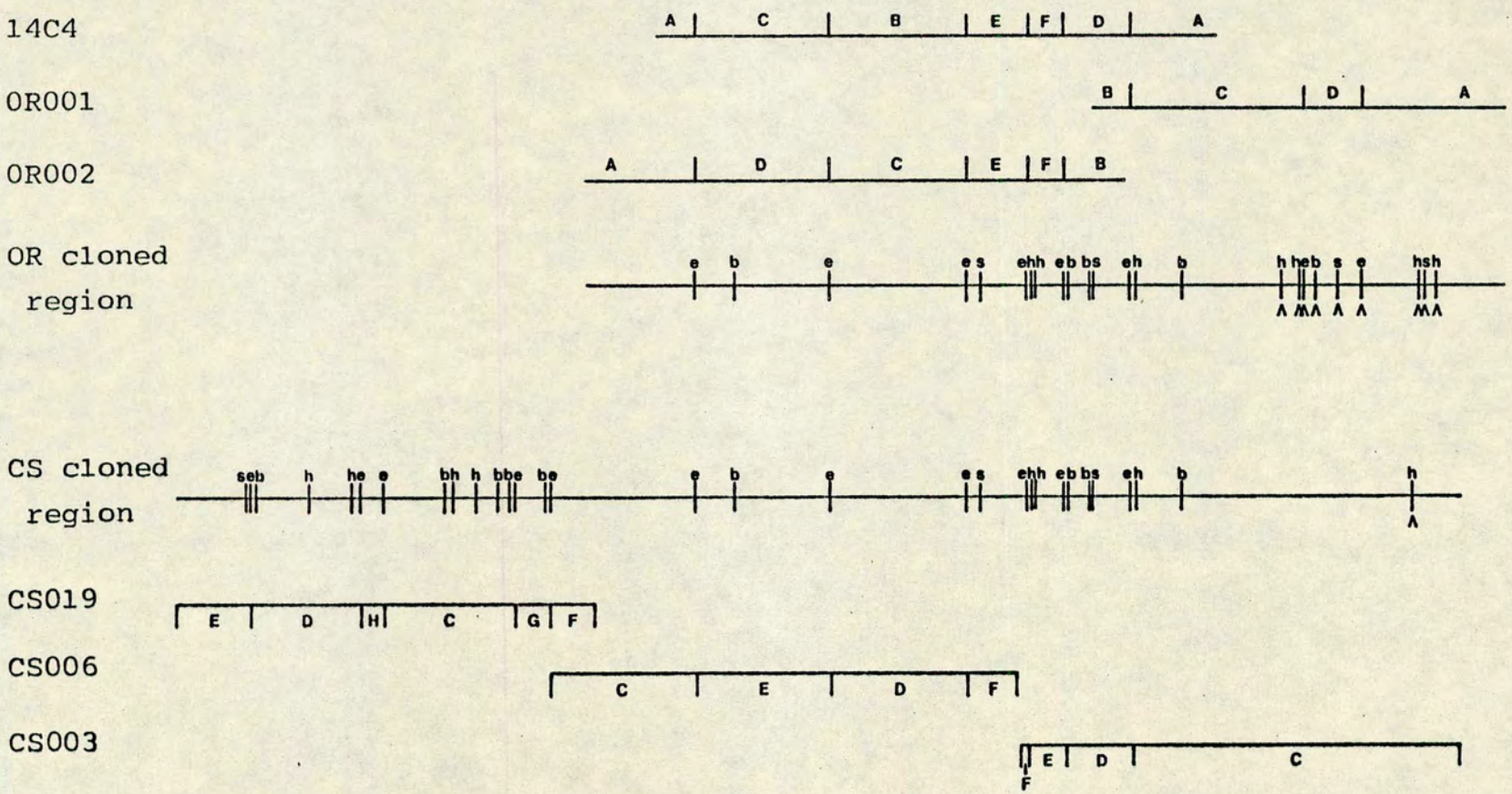

pDm2
A C B D A
OR006
A C E B
OR cloned region
e eh e e s b h h
CS cloned region
h bh s sbh h s h e eh e e s h he s hs e h s s
CS009
D G F H C E
CS007
B G F H I J K D E A
CS005
E C D

Figure 4.36


14C4
A C B E F D A
OR001
B C D A
OR002
A D C E F B
OR cloned region
e b e e s ehh eb bs eh b h heb s e hsh
CS cloned region
seb h he e bh h bbe be e b e e s ehh eb bs eh b h
CS019
E D H C G F
CS006
C E D F
CS003
F E D C

Discussion.

The results shown in figures 4.41 to 4.44 will now be considered. Additional copies of figures 4.35 and 4.36 are provided for ease of reference. In all cases (except that of CS003, where the reasons will become apparent) those fragments derived entirely from within the recombinant are also found in the embryo DNA tracks (unless of course they were too small to appear on the autoradiograph). Thus the arrangement of sequences found on these recombinants must occur within the Oregon R genome. If any of the recombinants carries a double insert, the breakpoint must lie within the terminal fragment.

The intensities of the internally derived fragments in the recombinant and embryo DNA tracks shown in figure 4.43 are approximately equal. Thus it can be concluded that (1) the sequence organisation found on these fragments (of CS005, CS006, and CS007) must be that most commonly found in Oregon R strain embryos, and (2) that the sequences represented by these fragments are not repeated intact. By logical extension of this second conclusion, none of the internal fragments observed in figures 4.41, 4.42 or 4.44 are repeated intact either.

Turning to the other fragments observed in the embryo DNA tracks, it is clear both from the number of bands observed and the overall intensity of the tracks, that CS003 and CS007 carry repetitive sequence. There are a few extra bands in most of the other tracks; if these are due to repeated sequences, then there can be no more than one or two copies of the repeat.

The results will now be analysed in detail. It should be apparent from preceding discussion that there are many possible explanations for a small number of extra bands. The interpretations given below are the simplest; they are certainly not the only ones.

pDm2.

Bam H1 - The single strong band suggests that the Bam H1 site found in pDm2 is uncommon in Oregon R embryos. The weakly labelled bands of 9 and 7kb could represent the expected two fragments present in a minority of the population. However, the Bam H1 site is found in both pDm2 and OR006, strongly suggesting that it is a common feature at this site (although it is absent from CS009). In addition the fragments are all too short to correspond to the positions of the adjacent Bam H1 sites as indicated by CS005 and CS007. The expected sizes would be 33.35+kb without the Bam H1 site, and 19.0 and 14.35+kb with the Bam H1 site. A more tenable hypothesis is that the 15+kb band consists of two fragments of almost identical size; these would be the 19.0 and 14.35+kb fragments. The two faint bands could be due to any of the reasons set out above, except partial digestion.

Eco R1 - The fragment extending to the left is most likely to be the weakly labelled 15+kb band. The fragment extending to the right is more difficult. The only band of sufficient intensity is the 5.5kb one. This is in excellent agreement with the expected size from the Canton S recombinants. Alternatively, the expected intensity could be made up from two or three other bands, indicating variation in this direction. The former is thought more likely. In OR006 there is an Eco R1 site missing relative to CS005 and CS009; the expected size is 7.5+kb, 12.85kb if the next site to the right in CS005 is present. This could be the other 15+kb band, which, being labelled to lower intensity than the 5.5kb band, would indicate that the Eco R1 site is more often present than absent in Oregon R embryos. The other bands must be due to one or other of the reasons mentions above.

This experiment was performed independently by G. Miklos using a different stock of Oregon R flies as his source of DNA. He obtained the same set of bands as shown in figure 4.41, but at different intensities, the difference

being such that the extra bands are unlikely to be partials (personal communication). This supports the suggestion of heterogeneity at this site. Further, although partials of roughly the right size could be produced, it is extremely unlikely that they would be of the relative intensities found here.

Hind 111 - The 4 and 1.5kb fragments are in excellent agreement with those expected. The additional fragment of about 5kb could be a partial, given the error in determining fragment length.

Pst 1 - The fragment extending to the left has such a short homology with pDm2 that it is unlikely to be detectable. It is just possible however that this is the 3kb band. The 8.5kb fragment must extend to the right on the basis of intensity. Thus there is at least one fragment unaccounted for.

Sma 1 - The two strongly labelled bands are in excellent agreement with expectations. The 15+kb band could be a partial.

Conclusions.

(1) The Eco R1 site found in CS005 and CS009, but missing in OR006, is probably more commonly present than absent in Oregon R embryos. Thus the absence of this site in OR006 represents variation in sequence within the population, rather than a strain specific difference.

(2) The possibility of an alternative arrangement of the sequences carried by pDm2, or of a short low frequency repeat, cannot be excluded.

14C4.

There is a repeated sequence present on 14C4, copies of which are found on CS016, CS020, and OR028 (see section 4.3). Additional bands derived from the regions of the genome represented by these recombinants, and from other copies of the repeat, may be observed, although the length of homology is close to the minimum detectable in these experiments.

Bam H1 - The fragment most likely to extend to the left of 14C4 is the 5.7kb one; this is in good agreement with that expected. The 3.7kb band fits with that predicted to lie to the right from the map of OR001. No fragment is found which corresponds to the 7.7+kb fragment expected from the map of CS003, suggesting that the arrangement of sequence found on CS003 is rare in Oregon R embryos, presuming it to occur at all. From the maps of CS016, CS020, and OR028, bands of 14.35+ and 13.6+kb would be expected; these are not present. Two fragments cannot be accounted for.

Eco R1 - The 4.2kb fragment corresponds to that expected to lie to the left of 14C4, and the 5.0kb to that lying to the right based on OR001. The 15+kb band could be that corresponding to the sequence arrangement found on CS003; if so, this arrangement occurs in only a minority of Oregon R embryos. The sizes of the fragments expected from the maps of CS016, CS020, and OR028 are 6.45, 5.85, and 2.8kb; these may be the 6.5, 5.7, and 2.7kb fragments, although their relative intensities are not those anticipated from the cross-hybridisation experiments. This discrepancy would be explained if these particular patterns of restriction sites were only found in a proportion of individuals. The 3.2kb fragment may originate from another copy of the repeat.

Hind 111 - The 15+kb fragment extends to the left of 14C4, and is in good agreement with the expected size. The 4.3kb band is that predicted from the map of OR001 to lie to the right. The 7.9kb band expected from the map

of CS003 might well be that of 8.5kb; again, if this is so, CS003 represents a minority arrangement in Oregon R embryos. The Hind 111 maps of CS016 and CS020 are not available, but fragments of 5.05, 4.55+, 1.45, and/or 0.75kb would be expected from OR028, depending on the position of the repeated sequence(s) within the C fragment of OR028. Thus the 1.5kb fragment could also be accounted for.

Pst 1 - The 4.5 and 2.8kb are not on their own sufficient to be compatible with the map of 14C4. The 2.8kb fragment is too short to lie to the right of 14C4. The intensity of hybridisation is roughly half that of the next largest band, which is an internal fragment of 3.75kb. Thus the region of homology on the 2.8kb fragment must be about 2kb, which is in reasonable agreement with that expected for a fragment lying to the left. The 4.5kb band is far too weak to extend in either direction in the majority of embryos. One can only suggest that the upper band on the autoradiograph, taken to be an internal fragment, is in fact a doublet. It is not possible to say whether this is so or not on the basis of intensity of hybridisation. Supposing this to be the case, this extra fragment would lie to the right. The 4.5kb fragment could now correspond to the minority sequence arrangement found in CS003, or to a copy of the repeat. If the former were true, the break in homology between CS003 and OR001 would be within 1.4kb of the end of 14C4. This is of course purely speculation; many other explanations are possible.

Sma 1 - Neglecting the levels of hybridisation, one would have said that the 15+kb band corresponds to the fragment extending to the left of 14C4, and the 7 and 10kb bands to the fragments predicted from the maps of OR001 and CS003 respectively, with the sequence arrangement found on CS003 being considerably less common. However were this the case, the 15+kb band would be expected to be more, and the 7kb band less intensely labelled. It is possible that

this discrepancy could be due to inefficient transfer of the larger fragment. Fragments of 8.6, 6.85+, 4.5, and/or 2.75+kb would be expected from CS016, CS020, and OR028. The observed fragment of 4.8kb could thus be accounted for.

Conclusions.

(1) The sequence arrangement found on CS003 is at best uncommon, and may possibly not occur, in Oregon R strain embryos.

(2) The repeated sequences present on 14C4, copies of which are found on CS016, CS020, and OR028, are too short to be reliably detected by this method under the conditions used.

(3) The possibility of an alternative arrangement of the sequences carried by 14C4 cannot be excluded.

CS003.

The insert of CS003 clearly includes a sequence or sequences which are repeated many times in both Canton S and Oregon R strain embryo DNA. However, the pattern of fragments obtained with the different DNAs is not the same; both the positions and relative intensities of the bands differ (see figure 4.44). This is most readily interpreted as indicating that the copies of the repeat occur (in some cases) at different sites in the genomes of the two strains.

From figure 4.43 it can be seen that there is an Eco R1-Hind 111 fragment of about 2.2kb which is labelled to high intensity. This suggests that in many cases the sequences present on this fragment are repeated intact (at least within Oregon R embryo DNA). No such fragment is found within the insert of CS003, although it could extend beyond the right end. The fragments predicted from the maps shown in figure 4.36 may or may not be present; the level of hybridisation precludes their identification.

CS005.

The 3.7kb fragment probably extends to the left of CS005, and the 2.8kb to the right. This leaves two, or perhaps three, fragments unaccounted for. Fragments of this size could be generated by partial digestion, but the possibility of low frequency repeats, alternative arrangements of sequences, or heterogeneity of restriction sites cannot be excluded.

CS006.

The 1.7kb band is in perfect agreement with that predicted to extend to the right of CS006. The expected 4.2kb band is not observed; the most intensely labelled new fragment is that of 4.1kb. The level of hybridisation to this fragment is slightly, but significantly, less than would be expected by comparison with adjacent bands. Note that seven of the recombinants isolated in the second

screening experiment contain a 4.2kb Eco R1 fragment, and none at 4.1kb Eco R1 fragment. Thus the simplest explanations are that either there is a Hind 111 site just within the 4.2kb Eco R1 fragment in the majority of cases, and that this is missing in CS006, or that the 4.1kb fragment is specific to Oregon R strain embryos, and the 4.2kb to Canton S.

The Hind 111 sites of CS016 and CS020 have not been mapped. Digestion of OR028 DNA with Eco R1 and Hind 111 would generate fragments of 2.45, 1.8, 1.45, and 0.75, some or all of which, depending on the exact arrangement of the repeated sequence, might be expected to hybridise. The 2.4kb band might correspond to the largest of these. The remaining bands could be due to any of the reasons set out above.

CS007.

The insert of CS007 clearly carries a sequence or sequences which are repeated in Oregon R strain embryo DNA. The expected new bands may or may not be present, being obscured by the hybridisation to copies of the repeat. The intensity of hybridisation to the extra bands is low, indicating that the repetitive sequence or sequences are very short, or poorly conserved. If the copies were of high fidelity and occurred at the same sites in all individuals, the repeats could be no longer than about 1kb.

OR001.

Note first that the tracks digested with differing amounts of Eco R1 are identical; none of the bands are due to partial digestion (see section 1).

Consider the Oregon R embryo DNA tracks. The two bands of 3.0 and 2.7kb are of odd appearance, and will be assumed to be artifacts. The 1.9kb band corresponds exactly with that expected to extend to the left of OR001 (as drawn in figure 4.36). The 5.7kb band most likely extends to the right, but is of slightly lower

density than expected. The 8.5kb fragment could be an alternative to the 5.7kb, thus accounting for the missing intensity, but other explanations are possible. One fragment remains unaccounted for. There is no sign of a fragment corresponding to the 9.3+kb expected from the map of CS003, indicating that if such an arrangement occurs in Oregon R embryos, it is uncommon.

Now consider the Canton S embryo DNA tracks. Again the 1.9kb band is found, suggesting that the sequences at this end of OR001 are invariant. The 5.7kb fragment is again present, but at much reduced intensity. The 8.5kb fragment is correspondingly more intense. This supports the suggestion made above that these two fragments represent alternatives to the right end of OR001, and indicates that the proportion of the two alternatives differs in the two strains. Two new fragments are found in the Canton S embryo DNA, indicating that there is an arrangement of sequence found in this strain which occurs infrequently, if at all, in Oregon R embryos. The 10kb fragment could correspond to the 9.3+kb fragment expected from the map of CS003. Now if CS003 represents a genuine alternative arrangement of sequence to that found on OR001, then the intensity of the internal 4.9kb fragment of OR001 should be reduced, since the breakpoint lies in here. It is difficult to say whether this is so or not, but the possibility certainly cannot be excluded. Any reduction in intensity is slight; the arrangement found on CS003 must be uncommon. If we accept this hypothesis, two alternatives exist: that the sequences present on CS003 but not on OR001 are inserted into the sequences found on OR001; that the sequences present on CS003 but not on OR001 replace the right end of OR001. In the former case, another fragment would be expected if the inserted sequence were longer than that found in CS003, and included at least one Eco R1 site. An additional fragment would not be expected if the right end of OR001 were replaced by the sequence found on CS003. A fragment

of 2.7kb is found. The level of hybridisation is somewhat less than that of the 10kb fragment. If this 2.7kb fragment were indeed due to insertion of the sequence found in CS003 into that found in OR001, this would place the point of insertion approximately 3kb from the left end of the 4.9kb fragment of OR001, set a minimum length for the inserted sequence of about 8kb, and locate an Eco R1 site within the inserted sequence about 1kb beyond the right end of CS003 (this situation is illustrated in figure 4.51). However it is also possible that the 2.7kb fragment is derived in some other way, in which case it represents a further strain specific difference, and that the sequence present on CS003 is inserted further to the right (and is even less common), any second fragment being too faint to be detected, or replaces the right end of OR001.

The 3.5kb fragment found in the Oregon R DNA track is not detected in Canton S. The failure to detect this fragment in Canton S strain DNA may be due to the slightly lower amount of DNA in this track, or represent a strain specific difference.

4.5 Localisation of Repetitive Sequences.

DNA from recombinants pDm2, 14C4, CS003, CS005, CS006, CS007, CS009, CS016, CS019, CS020, CS021, OR001, OR002, OR006, and OR028 was digested with Eco R1 (except in the cases of CS007 and OR006 where Hind 111 was used owing to the better distribution of sites), electrophoresed through 0.7% agarose gels and transferred to nitrocellulose. These filters were then probed with nick translated total genomic DNA from either Canton S or Oregon R strain embryos (see figures 4.45 and 4.46 respectively). The repetitive sequences carried by CS003 and CS007 were located with greater precision using additional restriction enzymes (see figures 4.47 and 4.48).

This type of experiment allows the identification of restriction fragments which carry repetitive sequences (see section 1). A filter bound DNA fragment consisting entirely of unique sequence will be labelled to an intensity proportional to its length by that unique sequence in the probe DNA. A fragment of the same length, but carrying a copy of a repeated sequence will be labelled more intensely, repeated sequences being present in higher concentration than unique sequences in the probe DNA. The increase in hybridisation will be related to the length, fidelity, and number of copies of the repeated sequences. Obviously the sensitivity of the technique depends on the length of the filter bound restriction fragment; the shorter the fragment, the greater the sensitivity.

Figure 4.45

DNA from recombinants pDm2, 14C4, CS003, CS005, CS006 CS007, CS009, CS016, CS019, CS020, CS021, OR001, OR002, OR006, and OR028 digested with Eco R1 (except CS007 and OR006, digested with Hind 111), electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated total genomic DNA from Canton S embryos: gel photograph (left) and autoradiograph (right). Note the intensity of hybridisation to the C fragment of CS003, the G fragment of CS007, the E fragment of CS021, and the C and E fragments of OR028.

Track	Recombinant
A	pDm2
B	14C4
C	CS003
D	CS005
E	CS006
F	CS007
G	CS009
H	CS016
I	CS019
J	CS020
K	CS021
L	OR001
M	OR002
N	OR006
O	OR028

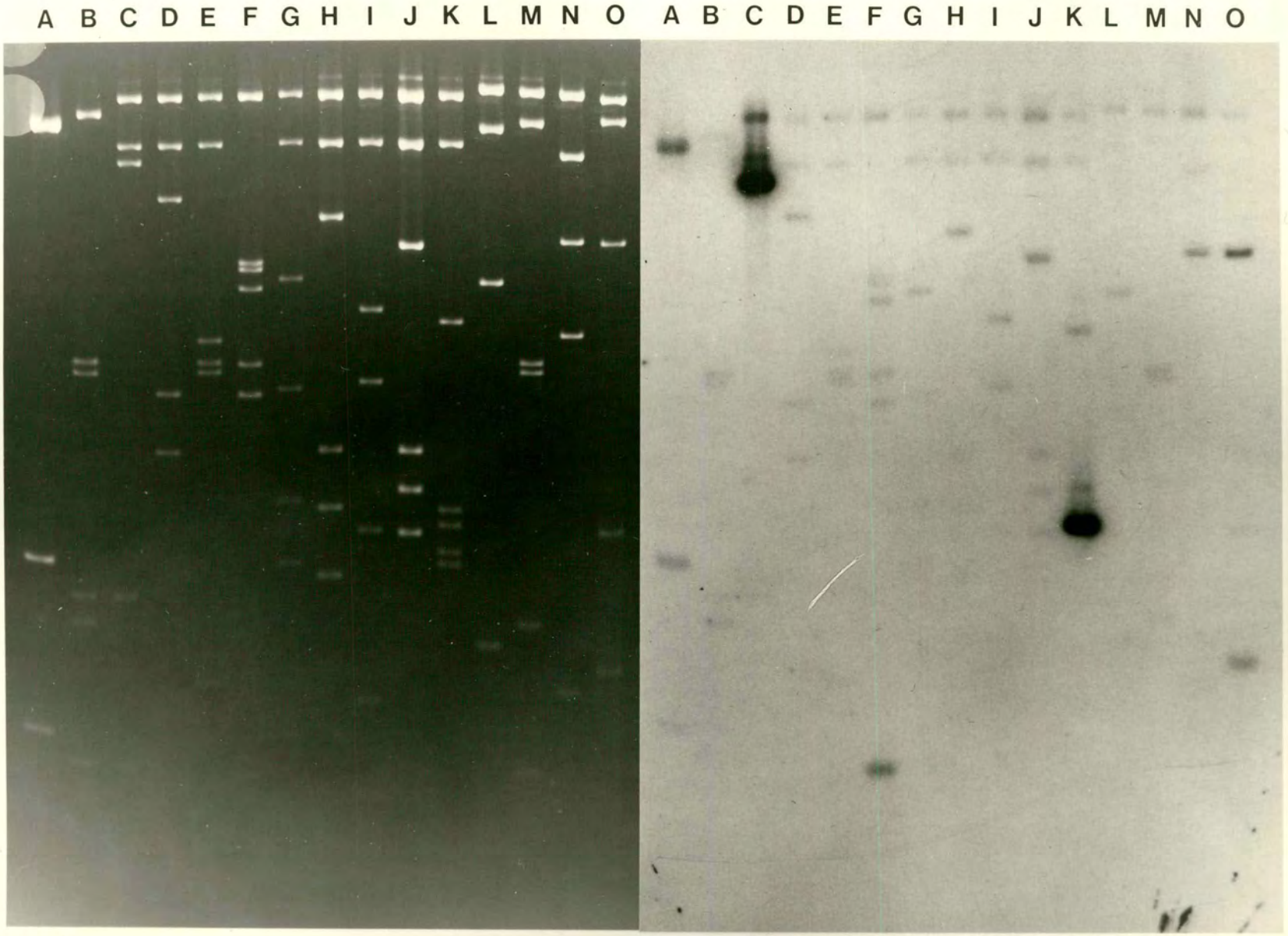
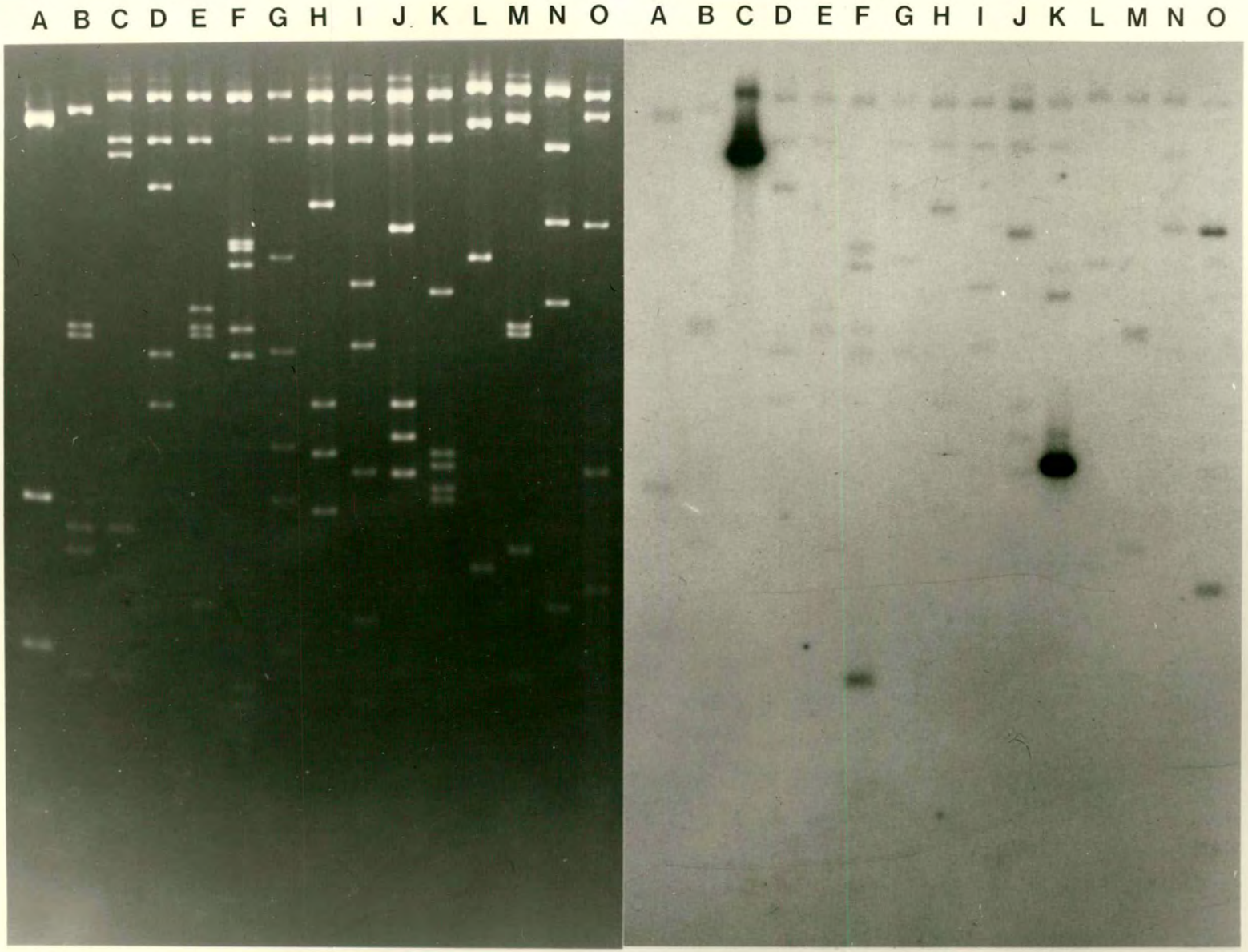

A B C D E F G H I J K L M N O
A B C D E F G H I J K L M N O

Figure 4.46

DNA from recombinants pDm2, 14C4, CS003, CS005, CS006 CS007, CS009, CS016, CS019, CS020, CS021, OR001, OR002, OR006, and OR028 digested with Eco R1 (except CS007 and OR006, digested with Hind 111), electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated total genomic DNA from Oregon R embryos: gel photograph (left) and autoradiograph (right). Note the intensity of hybridisation to the C fragment of CS003, the G fragment of CS007, the E fragment of CS021, and the C and E fragments of OR028.

Track	Recombinant
A	pDm2
B	14C4
C	CS003
D	CS005
E	CS006
F	CS007
G	CS009
H	CS016
I	CS019
J	CS020
K	CS021
L	OR001
M	OR002
N	OR006
O	OR028



4·46

Figure 4.47

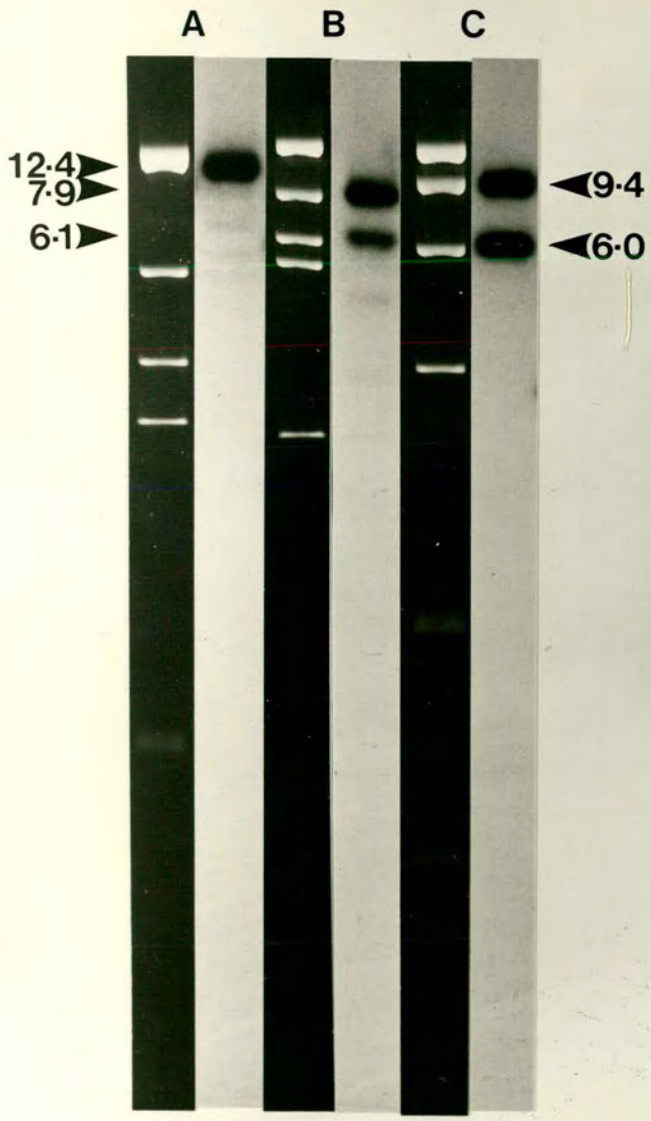
DNA from recombinant CS003 digested with Bam H1, Hind 111, or Sac 1, electrophoresed through a 0.7% agarose gel, and hybridised with nick translated total genomic DNA from Oregon R strain embryos: gel photograph (left) and autoradiograph (right). Arrowheads indicate fragments carrying repetitive sequences, giving their sizes in kb.

Track	Enzyme
A	Bam H1
B	Hind 111
C	Sac 1

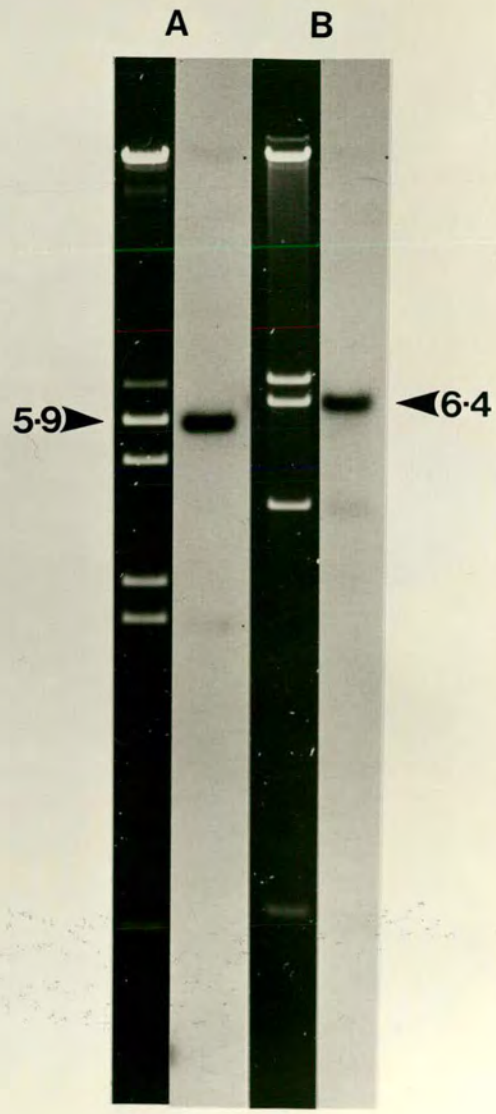
Figure 4.48

DNA from recombinant CS007 digested with Bam H1 and Sma 1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with nick translated total genomic DNA from Oregon R strain embryos: gel photograph (left) and autoradiograph (right). Arrowheads indicate fragments carrying repetitive sequences, giving their sizes in kb.

Track	Enzyme
A	Bam H1
B	Sma 1



4·47



4·48

Discussion.

Note first the level of non-specific hybridisation in figures 4.45 and 4.46 indicated by the labelling of vector DNA fragments. Fragments which carry unique sequence only will be labelled by virtue of both that unique sequence and this non-specific hybridisation. The longer the fragment, the stronger the labelling. Thus if no repeated sequences are present the level of hybridisation should be proportional to the strength of the band on the gel photograph. Only fragments which are clearly labelled above this background level will be regarded as containing repetitive elements.

There are five fragments which carry sequences repeated in both Canton S and Oregon R embryo DNA: the C fragment of CS003, the G fragment of CS007, the E fragment of CS021, and the C and E fragments of OR028. It is conceivable, but unlikely, that the sequences repeated are different in the two strains.

The repetitive regions of CS003 and CS007 were delineated more accurately in the experiments shown in figures 4.47 and 4.48 respectively. The repeat on CS003 lies to the right of the rightmost Bam H1 site (as expected from the failure to detect such repeated sequences on 14C4), and on both sides of the rightmost Hind 111 and Sac 1 sites. At this level of resolution the repetitive sequence appears to be continuous; it is possible that at finer resolution interspersion of short unique sequences would be found. The repetitive sequence begins in the restriction fragment which contains the discontinuity between CS003 and OR001. This is consistent with the repetitive sequence being inserted in some individuals of the Canton S strain into the region of the genome found uninterrupted in the C fragment of OR001, as suggested in section 4.4. This would represent another case where a sequence, repeated in both strains, is found at different sites in the two strains (see section 1).

The repetitive sequence found on CS007 is confined to

the region between the leftmost (as drawn in figure 4.35) Bam H1 and Hind 111 sites. It is quite possible that it extends to the left end of CS007, the additional labelling of the large B fragment going undetected against the background of vector cross-hybridisation.

It will be recalled that the C fragment of CS003, an unidentified fragment of CS007, and the E fragment of CS021 cross-hybridise, indicating that they carry copies of a repeated sequence (or sequences - see section 4.3). It seems likely that these repeated sequences are related to those detected in this series of experiments. If we accept this as being so, it is clear that the repetitive sequences found on CS003 and CS021 do not always occur together as a block; rather the unit of repeat is a segment of this block.

The C and E fragments of OR028 are either contiguous or separated by the 0.4kb G fragment which would not be present on the autoradiographs. Hence it is quite possible that the repetitive sequence is present as a single block.

In addition to these fragments, the A, B, and C fragments of pDm2 seem to be labelled more heavily with the Canton S embryo DNA probe than would be expected for unique sequence. This is not the case when Oregon R embryo DNA is used as probe. If this were a real phenomenon rather than an artifact, one would expect the appropriate fragments of CS005, CS007, CS009, and OR006 to show a similarly increased level of hybridisation with the Canton S DNA probe. This is indeed observed in the case of the E fragment of CS005, the C fragment of CS009, and the A and C fragments of OR006, although the effect is so slight that it might easily have been overlooked. Neither the E fragment of CS007, nor the G fragment of CS009 show labelling above background, although the latter is so faint that even a much larger increase in intensity would not be detected. These results are consistent with the repetition of most, if not all, of the sequences carried by pDm2, and possibly

part of the sequences which lie to the right, in Canton S but not Oregon R embryo DNA. The number of copies would have to be very low, say three at most. However, further experimentation is necessary before this conclusion can be regarded as other than extremely tentative. The simplest experiment would be to probe transfers of restricted Canton S and Oregon R embryo DNA with nick translated pDm2 DNA (with the appropriate controls against partial digestion), the prediction being that if the pDm2 sequences are repeated as described, more fragments will be observed in the Canton S than Oregon R DNA track.

Other than these, the fragments are labelled with the intensity expected for single copy DNA. The group (b) repeated sequences of 14C4, CS003, CS006, CS016, CS020, OR002, and OR028 identified by cross-hybridisation (see section 4.3) were not detected. However, if these are, as suggested, tRNA genes, they would have to be repeated at least thirty times before the increase in hybridisation to even the smallest fragment would be noticeable (see section 1). Similarly the sequence proposed to be repeated on the C fragment of CS016 and the E fragment of CS020 was not identified. Again this probably reflects the limitations of the technique.

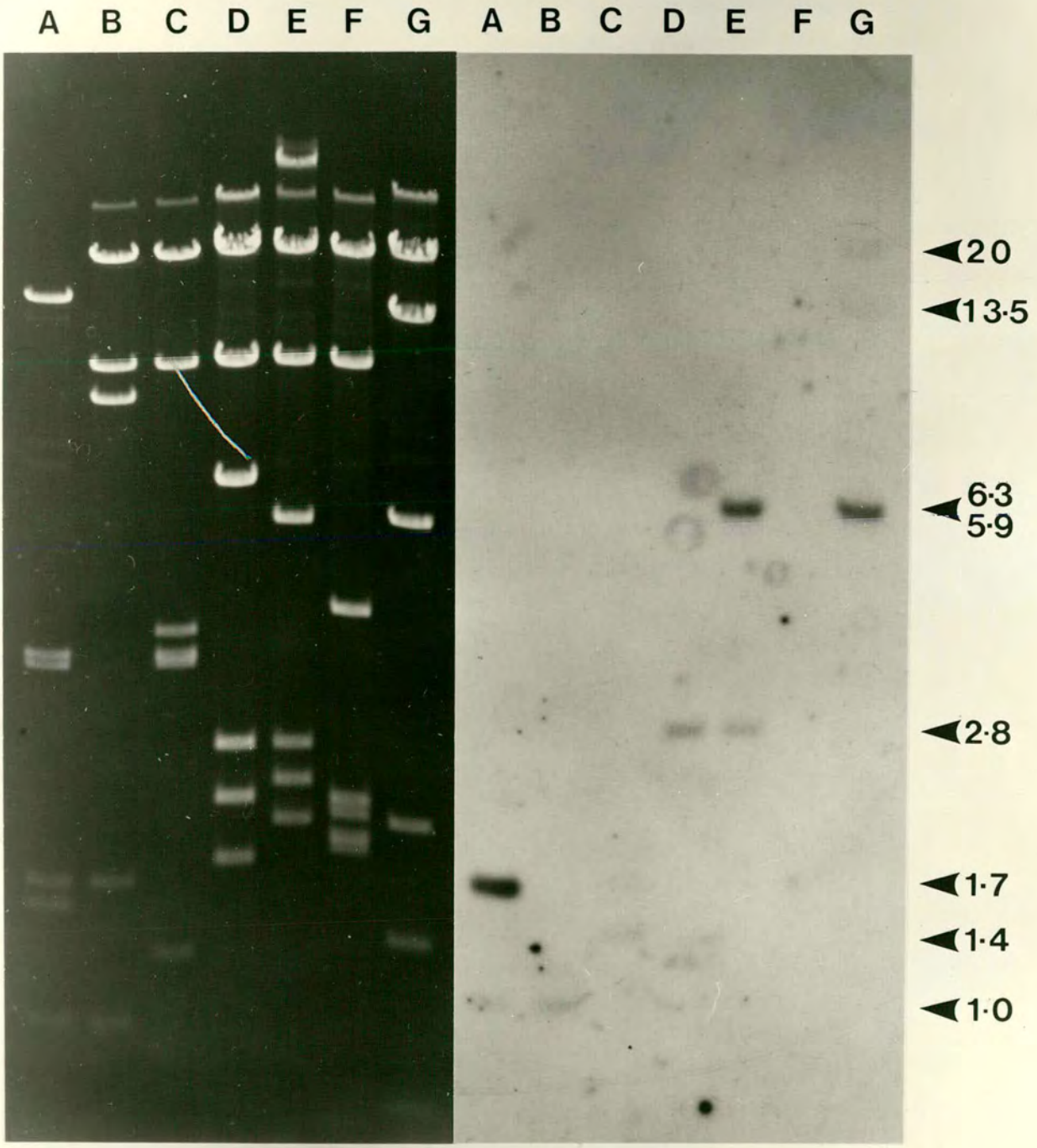
4.6 Localisation of tRNA genes.

There are short repeated sequences present on the E and F fragments of 14C4, the equivalent fragments of CS003, CS006, and OR002, the D fragment of CS016, the C and D fragments of CS020, and the C fragment of OR028 (see section 4.3). The E and F fragments of 14C4 contain tRNA genes (P. Schedl, personal communication); these might form part or all of the repeated sequences. In order to test this hypothesis DNA from recombinants 14C4, CS003, CS006, CS016, CS020, and OR028 was digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and probed with *in vivo* labelled 8 to 4S RNA from *Drosophila melanogaster* tissue culture cells (a gift from J. Dahlberg and E. Lund). The results are shown in figure 4.49.

Figure 4.49

DNA from recombinants 14C4, CS003, CS006, CS016, CS020, CS021 and OR028 digested with Eco R1, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridised with *in vivo* labelled 8 to 4S RNA from *Drosophila melanogaster* tissue culture cells (a gift from J. Dahlberg and E. Lund): gel photograph (left) and autoradiograph (right). Arrowheads indicate hybridising fragments giving their sizes in kb.

Track	Recombinant
A	14C4
B	CS003
C	CS006
D	CS016
E	CS020
F	CS021
G	OR028



4·49

Discussion.

It is clear that all the fragments of the group (b) recombinants identified as containing short repetitive elements by their behaviour on cross-hybridisation (see section 4.3) also carry sequences which have homology with 8 to 4S RNA. This strongly supports the suggestion made that at least part of the repeated sequences are tRNA genes, although other interpretations are possible. Note that in addition to these fragments, the A and B fragments of OR028 are labelled. In the cross-hybridisation experiments, any homology between these and the insert sequences of the other phage recombinants would have been lost in the hybridisation of vector DNA. However, no labelling was detected when 14C4 DNA was used as probe either.

The level of hybridisation observed is dependent on both the length of homologous sequence (or number of genes if all the hybridising RNA species are tRNAs) present on the filter bound fragments, and the abundance of the different hybridising RNAs in the probe. Therefore in the absence of additional information it is not possible to draw any useful conclusions from the relative intensities of the bands in figure 4.49.

The E and F fragments of 14C4 are both labelled with the 8 to 4S RNA probe. The E fragment of 14C4 is labelled much more intensely than the F fragment of CS006, indicating that the sequences which are present on the former, but lie beyond the right end of CS006 (see figure 4.36), are also transcribed. These data are consistent with there being a single tight cluster of tRNA genes in this region.

Both the C and D fragments of CS020 are labelled by the 8 to 4S RNA probe. The intervening G fragment is too small to appear on the autoradiograph. However, the E fragment of CS016, which contains those sequences of the C fragment of CS020 lying closest to the D fragment (see figure 4.37) is not labelled; therefore there must be two transcribed regions on CS020, separated by at least 2.45kb.

Similarly, the A, B, and C fragments of OR028 are not adjacent (see figure 4.32) and so there must be three discrete transcribed regions. Note that the low level of hybridisation to the A and B fragments may be an artifact; these fragments, being large, will be transferred with low efficiency.

4.7 Genomic occurrence of sequence organisation found on recombinants.

The degree to which the cloned sequences are representative of the organisation in the genomes of the two strains will now be considered.

All sequences which occur together on more than one recombinant can, with a very high degree of confidence, be taken to be representative of the genomic organisation of these sequences in a large fraction of the population (see section 1). In addition, if, on probing filters of restricted total genomic DNA with nick translated recombinant DNA, fragments of the sizes and intensities expected from restriction mapping of the recombinants are observed, this validates the sequence organisation found in those recombinants (see sections 1 and 4.4). Caution must be exercised when recombinants carrying inserts from different strains are compared, or when probing embryo DNA from one strain with a recombinant carrying an insert derived from another strain; a common arrangement of sequence in one strain may be a rare arrangement in another.

Thus from the results presented in sections 4.2, 4.3, and 4.4, it is clear that the arrangements of sequence found in pDm2, CS005, CS007, CS009, and OR006 are indeed co-linear in the genome, although the CS recombinants may represent an uncommon form of sequence organisation in the Canton S embryo genome.

There is some supporting evidence that the sequences carried by CS007 lie adjacent to those in CS009 in Canton S DNA. Preliminary characterisation of the products of the first screening experiment indicated that CS001, CS002, CS004, CS007, CS008, and CS011 all extend in the same direction. None of these recombinants contains Eco R1 sites other than those found on CS009 (see figure 4.3), suggesting that they carry the same sequences. This, by the way, would tend to indicate that the sequences carried by pDm2 are not repeated in

the Canton S genome (see section 4.5) in that, were this so, one would have anticipated isolating more than one sequence organisation extending in this direction. Ideally all recombinants should be mapped and used in extensive cross-hybridisation experiments, but in practice this would be too time consuming.

In conclusion, the pDm2 group of recombinants represent a naturally occurring sequence arrangement in both strains, this arrangement being common in Oregon R embryos, but possibly rare in Canton S. Note that the Eco R1 site found in CS005 and CS009, but missing in OR006 (see sections 4.2 and 4.4) seems usually to be present in Oregon R DNA.

Turning to the 14C4 group of recombinants, the sequences lying between the left end of CS006 and the point of discontinuity between CS003 and OR001 are validated by overlap and by the experiments described in section 4.4, although the CS recombinants could represent uncommon forms of organisation in the Canton S embryos. The possibility that there is in the Canton S genome an additional segment of DNA between the right and left ends of CS006 and CS003 respectively cannot be excluded. Also, as discussed in section 4.4, there may usually be a Hind 111 site just inside the C fragment of CS006, or some variation between the strains at this position. The only evidence in favour of the sequence organisation found on CS019 occurring normally within the genome is that from the probing of Hind 111 digested Oregon R embryo DNA with a 14C4 DNA probe; the fragment expected from the map of CS019 is observed. However, it must be added that the error in determining the size of fragments of this length is considerable. The situation to the right of 14C4 is more complex. The data suggests that the organisation found on CS003 does not occur in Oregon R embryo DNA, and is present only in a small minority of cases in the Canton S strain. The simplest interpretation is that in this small fraction of the Canton S embryo population, a repeated sequence of at

least 8kb is inserted into the region corresponding to the C fragment of OR001. There is additional evidence of variation in the position of the next Eco R1 site to the right of the end of OR001. This site would seem to occur either 5.7 or 8.5kb to the right of the rightmost Eco R1 site found in OR001 in both strains, the former being the more common in Oregon R embryos, and the latter in Canton S embryos.

The sequence arrangement found in the overlapping region of CS016 and CS020 must be common in the Canton S genome. Of the regions outside this overlap, nothing can be said, other than that double inserts are extremely unlikely in the CS library.

There is no evidence regarding the genomic occurrence of the arrangement of sequences found in OR028. Double inserts will not be present in the OR library, it being constructed by tailing.

4.8 Summary

The conclusions reached in the preceding sections are summarised in figures 4.50 (pDm2 group), 4.51 (14C4 group), 4.52 (CS016 and CS020) and 4.53 (OR028). In all cases the scale is 1cm = 2kb.

Figure 4.50

Diagram of the region of the Drosophila melanogaster genome represented by the pDm2 group of recombinants (pDm2, CS005, CS007, CS009, and OR006). Bam H1 (b), Eco R1 (e), Hind 111 (h), and Sma 1 (s) sites are shown. The sections of the cloned region carried by each individual recombinant are indicated. The boxed region contains repetitive sequences.

Notes

(1) The repetitive sequences may well extend to the left end of the cloned region. The arrangement of restriction sites is such that this could not be readily tested.

(2) The dotted region contains an additional three Hind 111 sites which could not be mapped.

(3) The Bam H1 site marked with an arrowhead is not found in CS009. This may indicate heterogeneity within either the Canton S or Oregon R strain, or a difference between the strains.

(4) The Eco R1 site marked with an arrowhead is not found in OR006. This appears to indicate heterogeneity within the Oregon R strain.

Figure 4.50

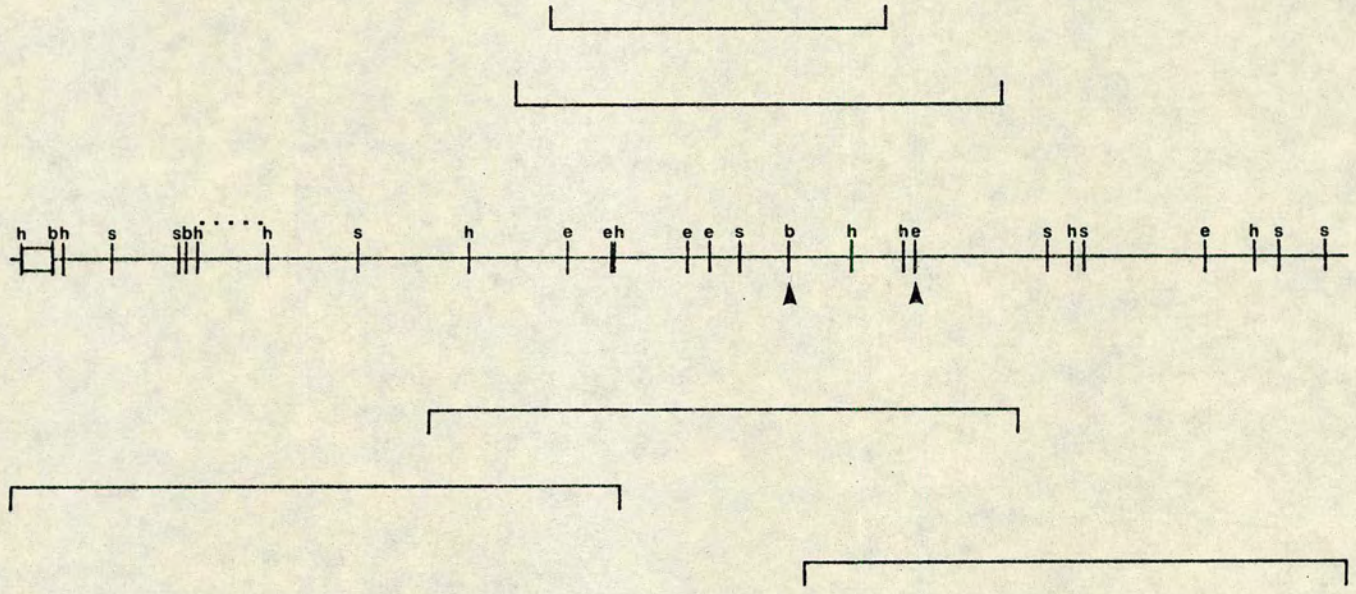

pDm2
OR006
Cloned region
h bh s sbh h s h e eh e e s b h he s hs e h s s
CS009
CS007
CS005

Figure 4.51

Diagram of the region of the Drosophila melanogaster genome represented by the 14C4 group of recombinants (14C4, CS003, CS006, CS019, OR001, and OR002). Bam H1 (b), Eco R1 (e), Hind 111 (h), and Sma 1 (s) sites are shown. The sections of the cloned region carried by each individual recombinant are indicated. The open boxed region contains repetitive sequences. The shaded boxed region contains sequences homologous to 8 to 4S RNA.

Notes

(1) The arrowhead marks the approximate site at which a moderately repetitive sequence, part of which is found in CS003, is believed to be inserted into the sequence found in OR001. The inserted sequence is found at this site in a proportion of Canton S embryo DNA, but not in Oregon R embryo DNA. The "looped out" regions show the proposed organisation of the inserted sequence and its alignment with CS003. The dotted region is conjectural; it may not be repetitive, and may include Bam H1, Hind 111, and Sma 1 sites. Additional Eco R1 sites may also be present, provided that the inserted sequence is correspondingly longer.

Figure 4.51

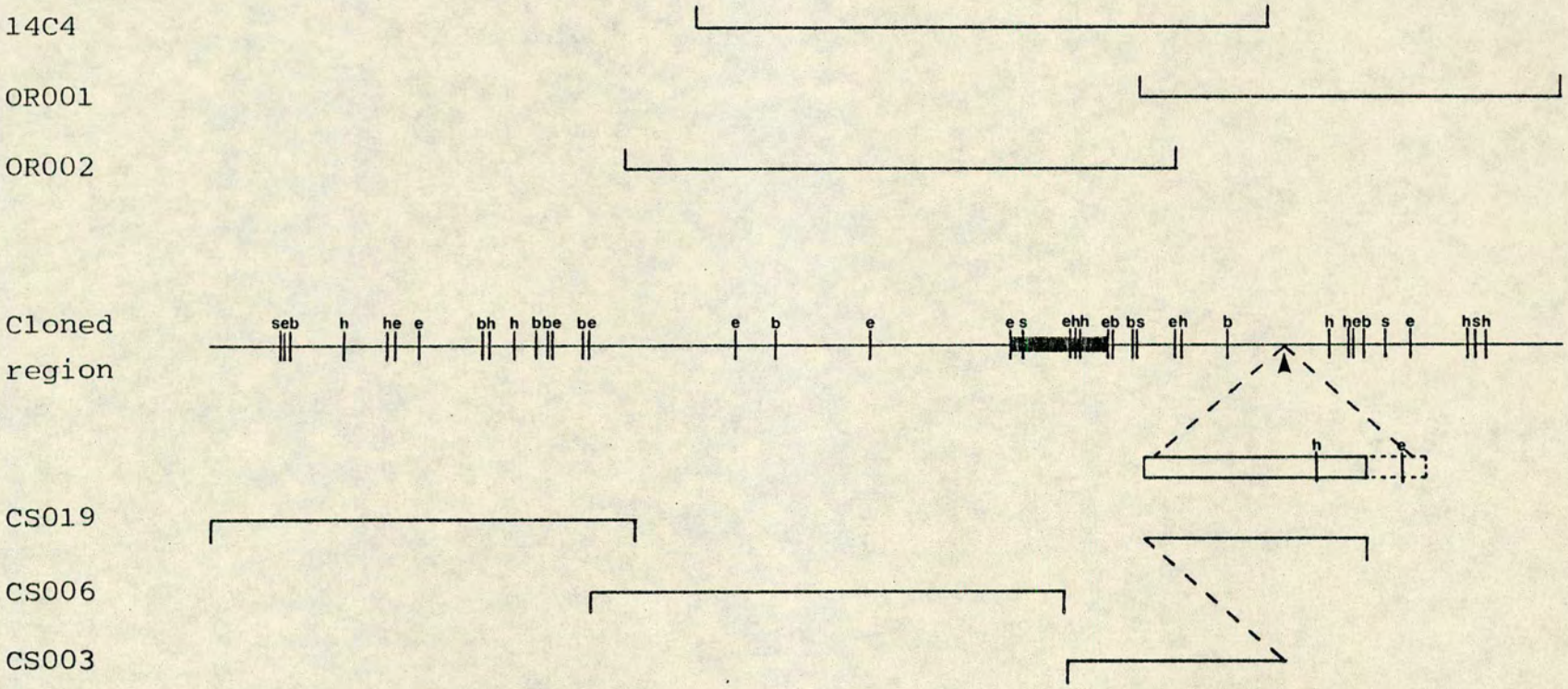

14C4
OR001
OR002
Cloned region
seb h he e bh h bbe be e b e es ehh eb bs eh b h heb s e hsh
h
e
CS019
CS006
CS003

Figure 4.52
Diagram of the region of the Drosophila melanogaster genome represented by CS016 and CS020. Bam H1 (b), Eco R1 (e), Hind 111 (h), Kpn 1 (k), Sac 1 (c), Sma 1 (s), and Xho 1 (x) sites are shown. The sections of the cloned region carried by each individual recombinant are indicated. The shaded boxed regions contain sequences homologous to 8 to 4S RNA. There are no detectably repetitive sequences.

Figure 4.52

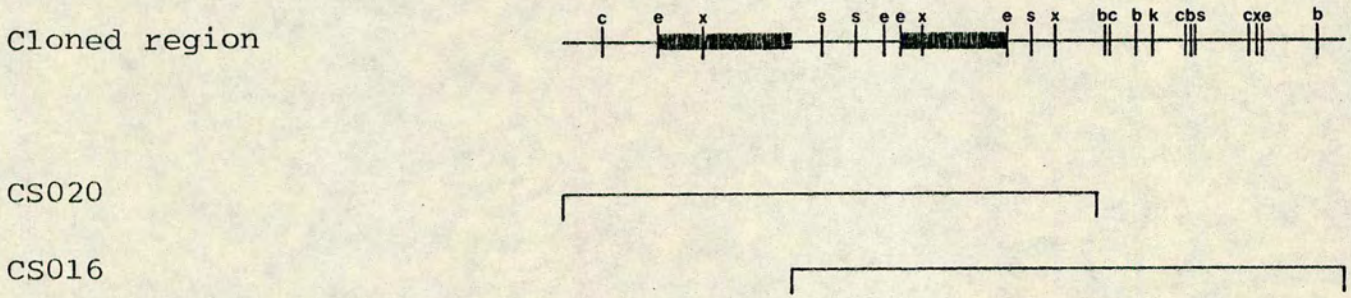


Figure 4.53

Diagram of the region of the Drosophila melanogaster genome represented by OR028. Bam H1 (b), Eco R1 (e), Hind 111 (h), Sac 1 (c), Sal 1 (1), Sma 1 (s), and Xho 1 (x) sites are shown. The open boxed region contains repetitive sequences. The shaded boxed region contains sequences homologous to 8 to 4S RNA.

Notes

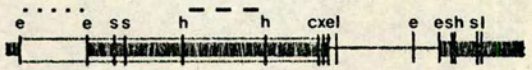
(1) The dotted region contains an additional Eco R1 site which could not be mapped.

(2) The dashed region contains an additional Hind 111 site which could not be mapped.

(3) The possibility that the repetitive sequence extends to the left end of the recombinant is not excluded by the data available.

Figure 4.53

OR028



Section 5.

DISCUSSION

5.1 Repetitive sequences

Regions of repetitive sequence were identified in the inserts of CS003, CS007, CS021, and OR028. The repetitive sequences proposed on the basis of cross-hybridisation intensities to occur in CS016 and CS020 were not detected on probing these recombinants with total genomic DNA and will not be considered further. The sequences hybridising 8 to 4S RNA will be dealt with in section 5.3.

The lengths of the repetitive sequences are not known; in all cases the possibility that they extend beyond the region represented by the insert cannot be excluded. The results are entirely consistent with the absence of the short period interspersion pattern of unique and repetitive sequences in Drosophila melanogaster (Manning et al, 1975; Crain et al, 1976; Lee et al, 1977). However the techniques used only indicate that particular restriction fragments carry repetitive sequence, and give no indication of unique sequences which might also be present. Hence it is possible that a more detailed analysis might reveal short unique sequences within the repetitive regions.

The inserts carried by CS003, CS007, and CS021 cross-hybridise, although there is no region of overlap. The fragments of CS003 and CS021 which are involved are those shown to carry repetitive sequences. The relative intensities observed when these three recombinants are cross-hybridised and probed with total genomic DNA strongly suggest that the repetitive regions are derived from scrambled clusters (Wensink et al, 1979) rather than from different copies of a 412-like family (Finnegan et al, 1977).

The repetitive sequence carried by CS003 is not present at the equivalent site in OR001. The results obtained on probing total genomic DNA from Canton S and Oregon R strain embryos with CS003 or OR001 DNA are best interpreted as follows: the sequence in question is repeated in both the Canton S and Oregon R strains; it is found at

this site in some, but not all, Canton S strain embryo DNA; it is not found at this site in Oregon R strain embryo DNA; when it occurs at this site in Canton S DNA, it is inserted into the sequences found on OR001, rather than substituted for them (although the data do not preclude small deletions). It seems to be a general principle that in Drosophila melanogaster dispersed middle repetitive sequences are never found at the same sites in two strains; this is the tenth example (W. Bender and P. Spierer, personal communication; B. Will and D. Finnegan, personal communication; Yen and Davidson, 1980; Nature News and Views 288 538 - 540, 1980; see also Young, 1979; Strobel et al, 1979).

This finding has certain implications for the function of these middle repetitive sequences. Clearly, if a repeated sequence is found adjacent to a particular unique sequence in one strain, but not in another, it can have no essential function which is directly related to that unique sequence. Thus middle repetitive DNA of the type so far studied in Drosophila melanogaster cannot have any direct role in the regulation of gene expression. (Note however that it is difficult to propose a satisfactory model for coordinate regulation of gene expression in eukaryotes which does not involve repeated sequences. Davidson and Britten (1979) have suggested that repetitive sequences involved in regulation might have insufficient homology to be detected under the conditions generally used). The moderately repetitive DNA of Drosophila melanogaster is however reasonably conserved in sequence (Wensink, 1977). This suggests that there are some constraints upon its sequence, which generally implies some function. There are three ways round this. Firstly, moderately repetitive sequences may have arisen relatively recently in Drosophila melanogaster, such that they have not had time to diverge (Wensink, 1977). Alternatively, the sequences may be subject to some correction mechanism by which the different copies are periodically compared and

any variation eliminated. Lastly, it has been proposed that some sequences, which have no function in regard to the whole organism, may be selected on the basis of their viability within the genome. Such sequences have been termed "selfish" (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). The ability to survive within the genome may impose certain limitation on sequence; dispersed moderately repetitive DNA in Drosophila melanogaster may be "selfish".

5.2 Sequence conservation

A total of twenty five restriction sites were identified in the regions cloned from both strains. Two sites were found which were present in one strain but not the other. Were these differences to be strain specific they would indicate a sequence divergence of approximately 1% between the two strains, in close agreement with the results of W. Bender and P. Spierer (personal communication). The results obtained on probing total genomic DNA from Oregon R strain embryos with pDm2 DNA suggest that the absence from OR006 of the Eco R1 site found within CS005 and CS009 represents heterogeneity within the Oregon R strain, rather than a difference between the strains. Several unexpected bands were observed in experiments of this type (see section 4.4); these could be taken to indicate further heterogeneity within the Oregon R strain (see in particular the discussion of the results obtained using OR001 DNA as probe). Thus it would appear that there may well be as much variation within the Oregon R strain as there is between the Oregon R and Canton S strains.

5.3 Transcribed sequences

Sequences homologous to an 8 to 4S RNA probe were located in 14C4 (and corresponding fragments of CS003, CS006, and OR002), CS016, CS020, and OR028.

Dudler et al (1980) have studied a cluster of tRNA genes in 14C4. Their restriction map of 14C4 differs

considerably from that given in figure 4.20. Figure 4.20 is believed to be correct because (1) the data were unambiguous and (2) the fragment order is confirmed by the overlaps with CS003, CS006, OR001, and OR002. Digestion of DNA prepared from a fresh sample of 14C4 obtained from W. Gehring confirmed the map shown in figure 4.20, indicating that the recombinant used by Dudler et al was the same as that used in this study. Closer examination of the published map reveals that the Eco R1 fragments are of the same size as those shown in figure 4.20, and that the maps are related by inversion of the central region, although the published version lacks one Bam H1 and two Hind 111 sites.

Given these discrepancies, Dudler et al find that 1.8 and 1.2kb Eco R1 fragments hybridise labelled 4S RNA, the ratio of intensity being approximately 4:1. These results are in good agreement with those shown in figure 4.49. Dudler et al propose that there are three arginine and two asparagine tRNA genes in 14C4. The asparagine tRNA gene contains sites for Bgl 11, Hpa 1, and Sma 1 (Hoveman et al, 1980; Yen and Davidson, 1980). Only one Sma 1 site was located within the region showing homology with 8 to 4S RNA. This suggests that there is only one asparagine tRNA gene in 14C4. The technique of Dudler et al (hybridisation of *in vivo* labelled tRNA to filter bound recombinant DNA, elution, and identification of the eluted tRNA by two dimensional gel electrophoresis) is liable to error in making estimates of gene number; different tRNAs may hybridise with different efficiencies.

The recombinants CS016, CS020, and OR028 also carry sequences homologous to the 8 to 4S RNA probe. During the writing of this thesis, Yen and Davidson (1980) published the results of a "walk" in the 42 A region of *Drosophila melanogaster*. This "walk" covers 94kb, the central 46kb of which contain a loose cluster of eight asparagine, five lysine, four arginine, and one isoleucine tRNA genes. Comparison of the restriction map of this region with the maps of CS016, CS020, and OR028 indicates that the

sequences carried by these three recombinants originate from the region covered in the "walk" (at positions -15 to -2, -9 to +4, and +6 to +19 respectively, see Yen and Davidson (1980), figure 2). The restriction maps of CS016 and CS020 are in excellent agreement with the results of Yen and Davidson. There are some small discrepancies in the map of the region represented by OR028, but these may well be explained by the greater resolution of their data.

According to Yen and Davidson, the D fragments of CS016 and CS020 contain one arginine tRNA gene, and the C fragment of CS020 one arginine and one lysine tRNA gene. In OR028, the A fragment contains a lysine tRNA gene, the B fragment, three asparagine tRNA genes, and the C fragment one arginine, three asparagine, and one lysine tRNA genes. The relative intensities of hybridisation obtained with the 8 to 4S RNA probe in figure 4.49 are not at first sight compatible with these conclusions, unless the different tRNA species were present at significantly different concentrations. However, the discrepancies can largely be ascribed to inefficient transfer of the large A and B fragments of OR028, and to two of the three asparagine tRNA genes in each of the B and C fragments of OR028 being in inverted repeat, forming a snap-back structure, and thus being largely unavailable for hybridisation.

Both the C and E fragments of OR028 were shown to contain repetitive sequences. Yen and Davidson detected repetitive sequence on the E fragment only. The reason for this discrepancy is not clear. Yen and Davidson examined recombinants carrying inserts from both the Canton S and Oregon R strains. Their results suggest that in the former there is a slightly repetitive sequence at this position; in the latter a more highly repetitive sequence of length 0.25kb is inserted. The recombinant OR028 is derived from the Oregon R strain. Yen and Davidson used Canton S pupal DNA in their experiments to detect repetitive sequences; I used Canton S and Oregon R

embryo DNA. The difference in our results may reflect differences between the two stocks of Canton S used, or may indicate that the repetitive sequence found on the C fragment of OR028 is repeated in embryo, but not pupal DNA. Further investigation is necessary.

5.4 Future course of the "walk"

An attempt is to be made to clone the Antennapedia locus using the inversion In(3R)Hu (Lindsley and Grell, 1968), which brings 84 F1 - 2, the site of origin of the 14C4 group of recombinants, adjacent to 84 B1 - 2, the site of Antennapedia. The breakpoint in 84 B1 - 2 results in the Humeral phenotype (Lindsley and Grell, 1968). Humeral is one of a cluster of homeotic genes found around Antennapedia (T. Kaufman, personal communication).

It is also hoped that detailed characterisation of the cloned region will proceed, and, in particular that any transcribed sequences and nuclease sensitive regions be located. Knowledge of the organisation of the genome should lead to a better understanding of its function (see section 1).

REFERENCES.

Appleyard, R.K. (1954)
Genetics 39 440-452.

Artavanis-Tsakonas, S., Schedl, P., Tschudi, C., Pirrotta, V., Steward, R., and Gehring, W.J. (1977)
Cell 12 1057-1067.

Ashburner, M., and Bonner, J.J. (1979)
Cell 17 241-254.

Benton, W.D., and Davis, R.W. (1977)
Science 196 180-182.

Benyajati, C., and Worcel, A. (1976)
Cell 9 393-407.

Blattner, F.R., Williams, B.G., Blechl, A.E., Thompson, K.D., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. (1977)
Science 196 161-169.

Boyer, H.W., and Roulland-Dussoix, D. (1969)
J. Mol. Biol. 41 459-472.

Boyer, H.W., Betlach, M., Bolivar, F., Rodriguez, R.L., Heyneker, H.L., Shine, J., and Goodman, H.M. (1977)
in "Recombinant Molecules: Impact on Science and Society" eds. Beers, R.F., and Basset, E.G. pp. 9-20 (Raven Press).

Bridges, C.B. (1935)
J. Hered. 26 60-64.

Britten, R.J., and Davidson, E.H. (1969)
Science 165 349-357.

Brown, D.D., Wensink, P. C., and Jordan, E. (1971)
Proc. Nat. Acad. Sci. USA 68 3175-3179.

Clarke, L., and Carbon, J. (1976)
Cell 9 91-99.

Clarke, L., and Carbon, J. (1980)
Nature 287 504-509.

Collins, J., and Hohn, B. (1978)
Proc. Nat. Acad. Sci. USA 75 4242-4246.

Cory, S., and Adams, J.M. (1977)
Cell 11 795-805.

Crain, W.R., Eden, F.C., Pearson, W.R., Davidson, E.H., and Britten, R.J. (1976)
Chromosoma 56 309-326.

Daneholt, B. (1975)
Cell 4 1-9

Davidson, E.H., and Britten, R.J. (1979)
Science 204 1052-1059.

Davis, R.W., Simon, M., and Davidson, N. (1971)
in "Methods in Enzymology" 21D eds. Grossman, L., and Moldave, K. pp. 413-428 (Academic Press).

Denell, R. (1973)
Genetics 75 279-297.

Denhardt, D.T. (1966)
Biochem. Biophys. Res. Commun. 23 641-646.

Doolittle, W.F., and Sapienza, C. (1980)
Nature 284 601-603.

Dudler, R., Egg, A.H., Kubli, E., Artavanis-Tsakonas, S., Gehring, W.J., Steward, R., and Schedl, P. (1980)
Nucl. Acids Res. 8 2921-2937.

Duncan, I.W., and Kaufman, T.C. (1975)
Genetics 80 733-752.

Felsenfeld, G. (1978)
Nature 271 115-122.

Finch, J.T., and Klug, A. (1976)
Proc. Nat. Acad. Sci. USA 73 1897-1901.

Finnegan, D.J., Rubin, G.M., Young, M.W., and Hogness, D.S. (1977)
Cold Spring Harbor Symp. Quant. Biol. 42 1053-1063.

Gehring, W.J. (1978)
Isr. J. Med. Sci. 14 295-304.

Georgiev, G.P. (1969)
J. Theor. Biol. 25 473-490.

Glover, D.M., and Hogness, D.S. (1977)
Cell 10 167-176.

Grunstein, M., and Hogness, D.S. (1975)
Proc. Nat. Acad. Sci. USA 72 3961-3965.

Hershey, N.D., Conrad, S.E., Sodja, A., Yen, P.H., Cohen, M., Davidson, N., Ilgen, C., and Carbon, J. (1977)
Cell 11 585-598.

Hicks, J., Strathern, J.N., and Klar, A.J.S. (1979)
Nature 282 478-483.

Hozumi, N., and Tonegawa, S. (1976)
Proc. Nat. Acad. Sci. USA 73 3628-3632.

Igo-Kemenes, T., and Zachau, H.G. (1977)
Cold Spring Harbor Symp. Quant. Biol. 42 109-118.

Judd, B.H., Shen, M.W., and Kaufman, T.C. (1972)
Genetics 71 139-156.

Kedes, L.H. (1979)
Annu. Rev. Biochem. 48 837-870.

Klug, A., Rhodes, D., Smith, J., Finch, J.T., and Thomas, J.O. (1980)
Nature 287 509-516.

Laskey, R.A., and Mills, A.D. (1977)
FEBS Lett. 82 314-316.

Lee, A.S., Britten, R.J., and Davidson, E.H. (1977)
Cold Spring Harbor Symp. Quant. Biol. 42 1065-1076.

Lefevre, G. (1976)
in "The Genetics and Biology of Drosophila" 1a eds. Ashburner, M., and Novitski, E. pp. 31-66
(Academic Press).

Lewin, B. (1974)
"Gene Expression" 2 (John Wiley and Sons).

Lindsley, D.L., and Grell, E.H. (1968)
"Genetic Variations of Drosophila melanogaster"
(Carnegie Inst. Washington Publ.).

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., and Efstratiadis, A. (1978)
Cell 15 687-701.

Manning, J.E., Schmid, C.W., and Davidson, N. (1975)
Cell 14 141-155.

Moran, L., Mirault, M., Tissieres, A., Lis, J., Schedl, P., Artavanis-Tsakonas, S., and Gehring, W.J. (1979)
Cell 17 1-8.

Morata, G., and Lawrence, P.A. (1977)
Nature 265 211-216.

Murray, N.E., Brammar, W.J., and Murray, K. (1977)
Mol. Gen. Genet. 150 53-67.

Orgel, L.E., and Crick, F.H.C. (1980)
Nature 284 604-607.

Palmiter, R.D., Mulvihill, E.R., McKnight, G.S., and Senear, A.W. (1977)
Cold Spring Harbor Symp. Quant. Biol. 42 639-647.

Pardue, M.L., Brown, D.D., and Birnstein, M.L. (1973)
Chromosoma 42 191-203.

Paul, J. (1972)
Nature 238 444-446.

Rigby, P.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977)
J. Mol. Biol. 113 237-251.

Rubin, G.M., Finnegan, D.J., and Hogness, D.S. (1976)
Prog. Nucleic Acid Res. Mol. Biol. 19 221-226.

Rudkin, G.T. (1972)
in "Results and Problems in Cell Diffrentiation" 4 ed. Beerman, W. pp. 59-85 (Springer-Verlag).

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977)
Nature 265 687-695.

Schachat, F., O'Connor, D.J., and Epstein, H.F. (1978)
Biochim. Biophys. Acta 520 688-692.

Sgaramella, V., and Khorana, H.G. (1972)
J. Mol. Biol. 72 493-502.

Shen, C.J., and Maniatis, T. (1980)
Cell 19 379-391.

Southern, E.M. (1975)
J. Mol.Biol. 98 503-517.

Stahl, F.W. (1979)
"Genetic Recombinantion" (W.H. Freeman and Co.).

Strobel, E., Dunsmuir, P., and Rubin, G.M. (1979)
Cell 17 429-439.

Vaughan, S. (1977)
Nature 269 86-87.

Weintraub, H., and Groudine, M. (1976)
Science 193 848-856.

Wellauer, P.K., and Reeder, R.H. (1975)
J. Mol. Biol. 94 151-161.

Wells, R., Royer, H.D., and Hollenberg, C.P. (1976)
Mol. Gen. Genet. 147 45-51.

Wensink, P.C., Finnegan, D.J., Donelson, J.E., and Hogness, D.S. (1974)
Cell 3 315-325.

Wensink, P.C. (1977)
Cold Spring Harbor Symp. Quant. Biol. 42 1033-1039.

Wensink, P.C., Tabata, S., and Pachl, C. (1979)
Cell 18 1231-1246.

Worcel, A., and Benyajati, C. (1977)
Cell 12 83-100.

Wu, C., Bingham, P.M., Livak, K.J., Holmgren, R., and Elgin, S.C.R. (1979a)
Cell 16 797-806.

Wu, C., Wong, Y., and Elgin, S.C.R. (1979b)
Cell 16 807-814.

Wu, C. (1980)
Nature 286 854-860.

Yen, P.H., and Davidson, N. (1980)
Cell 22 137-148.

Young, M.W. (1979)
Proc. Nat. Acad. Sci. USA 76 6274-6278.